

Objektive Tests:

Empirische Untersuchung zu Integration und
Kontrast direkter und indirekter
Messverfahren

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Fakultät für Philosophie, Kunst-, Geschichts- und Gesellschaftswissenschaften
der Universität Regensburg

vorgelegt von

Dipl. Soz. Dominik Weigand

aus Marburg/Lahn

im Januar 2018

Regensburg 2018

Gutachterin (Betreuerin): Prof. Dr. Melanie Walter-Rogg (Universität Regensburg)

Gutachter: Prof. Dr. Horst-Alfred Heinrich (Universität Passau)

"TO EXPLORE STRANGE NEW WORLDS, TO SEEK OUT NEW LIFE AND NEW CIVILIZATIONS, TO BOLDLY GO WHERE NO ONE HAS GONE BEFORE."

(Star Trek: The Next Generation)

Für meine Eltern und meine Schwester:

Gregor, Ute und Anna-Katrin

Danksagung

An dieser Stelle möchte ich allen danken, die mich bei der Verfassung der vorliegenden Dissertation und den vorausgegangenen Befragungen entweder direkt oder indirekt unterstützt haben und deren Einsatz ich keinesfalls als selbstverständlich erachte:

Meiner Doktormutter Frau Prof. Dr. Melanie Walter-Rogg danke ich für Ihre vielfältige fachliche und moralischen Betreuung und Beratung sowie die sehr wertvollen Rückmeldungen während der Entstehung dieser Studie. Außerdem bedanke ich mich für die Freiheit, die Sie mir bei der inhaltlichen und methodischen Konzeption meines Dissertationsprojektes ermöglicht hat, für Ihr Vertrauen in meine Kompetenz, mein Dissertationsprojekt erfolgreich abzuschließen, für das kollegiale Verhältnis zueinander und die vielen Erfahrungen, die ich in den vergangenen 6 Jahren an der Professur für Methoden der Politikwissenschaft machen konnte.

Herrn Prof. Dr. Horst-Alfred Heinrich danke ich für die positiven Erfahrungen im Rahmen des Kolloquiums an der Universität Passau und vor allem für seine Bereitschaft, als Co-Gutachter zur Verfügung zu stehen.

Dem derzeitigen Team der Professur für Methoden der Politikwissenschaft und Ehemaligen Romy Escher, Dr. Maximilian Grasl, Jasper Hamberg, Tassilo Heinrich, Maximilian Hösl, Ann-Kathrin Reinl und Michaela Schmid danke ich für die Möglichkeit, in Ihren Lehrveranstaltungen Erhebungen durchführen zu können und die damit verbundene tatkräftige Unterstützung bei der Durchführung der Befragungen, die vielen Bürgesprache über Methodik, Politik und Anderes und das freundschaftliche Verhältnis miteinander.

Herrn Prof. Dr. Stephan Bierling, Herrn Prof. Dr. Martin Sebaldt und Herrn PD Dr. Alexander Straßner danke ich für die Möglichkeit, in Ihren Lehrveranstaltungen Befragungen durchführen zu dürfen.

Danken möchte ich außerdem allen weiteren Mitarbeitern des Instituts für Politikwissenschaft der Universität Regensburg, die in den vergangenen 6 Jahren meine Weggefährten waren.

Ebenso gilt mein Dank allen Studierenden, die an den Befragungen im Rahmen der Pilot- und Hauptstudie teilgenommen haben. Ohne die aus diesen Erhebungen gewonnenen Daten hätten die dieser Arbeit zugrundeliegenden empirischen Analysen nicht umgesetzt werden können.

Besonderer Dank gilt meiner Familie, ohne deren emotionalen Rückhalt und auch praktischen Einsatz der Abschluss meines Dissertationsprojektes nicht möglich gewesen wäre: Meinem Vater für das Korrekturlesen der Doktorarbeit und meiner Mutter und meiner Schwester für Ihren Beistand und die aufbauenden Gespräche.

Regensburg, im Januar 2018

Dipl. Soz. Dominik Weigand

Inhalt

1. Einleitung	1
2. Theorie: Methodik	7
2.1 Reaktivität	7
2.2 Direkte und indirekte Messverfahren	11
2.2.1 Direkte, explizite bzw. reaktive Messverfahren	14
2.2.1.1 Kognitive Reaktivitätseffekte	16
2.2.1.2 Soziale Reaktivitätseffekte	18
2.2.1.2.1 Soziale Erwünschtheit	19
2.2.1.2.2 Kontrolle sozialerwünschten Antwortverhaltens	21
2.2.1.3 Repräsentativität	23
2.2.2 Indirekte Messverfahren	23
2.2.2.1 Physiologische bzw. psychophysiologische Messungen	24
2.2.2.2 Nicht-reaktive bzw. unaufdringliche Messverfahren	25
2.2.2.3 Implizite Messverfahren bzw. Maße	30
2.2.2.4 Projektive Verfahren	35
2.2.2.5 Objektive Tests	38
2.2.2.5.1 Begriffsbestimmung	38
2.2.2.5.2 Herleitung und Definition	40
2.2.2.5.3 Zur Objektivität von Objektiven Tests	45
2.2.2.5.4 Zur Verfälschungssicherheit von Objektiven Tests	45
2.2.2.5.5 Zur Validität von Objektiven Tests	51
2.2.2.5.6 Probleme von Objektiven Tests	56
2.3 Fazit und Ausblick	58
3. Theorie: Vertrauen	61
3.1 Soziales Vertrauen	62
3.1.1 Ausdifferenzierung Sozialen Vertrauens	69
3.1.1.1 <i>Normative</i> oder <i>moral trust</i>	69
3.1.1.2 <i>Emotional</i> oder <i>affective trust</i>	74
3.1.1.3 <i>Identity</i> oder <i>group-based trust</i>	77
3.1.1.4 <i>Strategic</i> oder <i>rational trust</i>	80
3.1.2 Definition Sozialen Vertrauens	85
3.2 Generalisiertes Vertrauen	86

3.2.1 Determinanten <i>generalisierten Vertrauens</i>	86
3.2.2 Society-centered Accounts.....	91
3.2.2.1 Network-based Approach.....	91
3.2.2.1.1 Vertrauen und Soziales Kapital (Putnam)	92
3.2.2.1.2 Vertrauen als rationale Entscheidung (Herreros)	94
3.2.2.1.3 Thick und Thin Trust (Wollebæk/Selle)	95
3.2.2.1.4 Vertrauen aus sozialen Netzwerken (Stolle).....	96
3.2.2.2 Culture-based Approach	99
3.2.2.2.1 Vertrauen als Kontinuum mit zwei Polen (Uslaner).....	100
3.2.3 <i>Institution-centered Accounts</i>	104
3.2.3.1 <i>Attitudinal Approaches</i>	105
3.2.3.1.1 Staat und Vertrauen (Levi)	105
3.2.3.1.2 Vertrauen als Folge sozialer Ordnung (Offe)	107
3.2.3.1.3 Generalisierte Reziprozität (Newton)	113
3.2.3.2 <i>Institutional-structural Approach</i>	116
3.2.3.2.1 Vertrauensnormen als Folge von „ <i>street-level bureaucracy</i> “ (Rothstein/Stolle)	116
3.2.3.2.2 Institutionelle Regeln als Katalysator des Vertrauens (Freitag/Bühlmann).....	121
3.2.4 Ethnische Diversität.....	123
3.2.4.1 Der Einfluss statischer und dynamischer Diversität auf Vertrauen (Gundelach)	124
3.2.4.2 Similarität und Homogenität als Verstärker des Vertrauens (Öberg/Oskarsson/Svensson)	126
3.2.5 Generalisiertes Vertrauen vs. Partikulares Vertrauen	128
 4. Messung generalisierten Vertrauens	 132
4.1 Standarditems zur Messung <i>generalisierten Vertrauens</i>	132
4.2 Alternative Vertrauensskalen	136
4.2.1 Interpersonal Trust Scale (ITS)	136
4.2.2 Philosophies of Human Nature (PHN).....	139
4.2.3 Machiavellismus (Mach IV bzw. Mach V).....	140
4.2.4 The Trust Inventory	141
4.2.5 Inventar zur Erfassung des interpersonellen Vertrauens (IIV)	142
4.2.6 KUSIV3	142
4.3 Methodische und inhaltliche Probleme der Standarditems.....	143
4.3.1 Existenz <i>generalisierten Vertrauens</i>	143
4.3.2 Radius <i>generalisierten Vertrauens</i>	145
4.3.3 Vertrauenslevel	149

4.3.4 Reaktive Effekte	151
4.3.5 Multidimensionalität der Standarditems	152
4.3.6 Differenziertheit der Vertrauensmessung	153
5. Hypothesen	155
5.1 Integration von direkten und indirekten Messverfahren	155
5.2 Homogenität und Reliabilität.....	162
5.3 Validität.....	163
5.4 Differenz von hybriden Messverfahren und direkten Messverfahren	168
5.4.1 Reaktive Effekte	168
5.4.1.1 Relevanz der Augenscheinvalidität.....	173
5.4.1.1.1 Relevanz der Augenscheinvalidität für die Messwerte von Standard- und Objektiv-Messung	173
5.4.1.1.2 Relevanz der Augenscheinvalidität für die Differenz zwischen Standard- und Objektiv-Messung	175
5.5 Radiusproblem.....	176
5.5.1 Radiusproblem und soziale Erwünschtheit	179
6. Pilotstudie	180
6.1 Methodik.....	180
6.1.1 Forschungsdesign	180
6.1.1.1 Erhebungskonzept	180
6.1.1.2 Stichprobe.....	181
6.1.1.3 Feldzugang.....	182
6.1.1.4 Operationalisierung	182
6.1.1.4.1 Glück und subjektives Wohlbefinden/Zufriedenheit.....	182
6.1.1.4.2 Generalisiertes Vertrauen.....	183
6.1.1.4.2.1 Optimismus.....	184
6.1.1.4.2.2 Subjektive Schicksalskontrolle/Selbstkontrolle (Locus of Control)	184
6.1.1.4.2.3 Gemeinsames Schicksal/Gemeinsame Werte/Ähnlichkeit	185
6.1.1.4.2.4 Misanthropie	186
6.1.1.4.2.5 Vertrauensbereitschaft, Hilfsbereitschaft und Fairness.....	187
6.1.1.4.2.6 Formelle/Informelle Netzwerke	188
6.1.1.4.2.7 Institutionen/Soziale Ordnung	189
6.1.1.4.2.8 Heterogenität/Diversität	191
6.1.1.4.3 Kognitiver Pretest.....	191

6.1.1.4.4 Radiusproblem	192
6.1.1.4.5 Soziale Erwünschtheit	193
6.1.1.4.6 Sozioökonomischer Status	194
6.1.1.4.7 Mobilität und soziale Gemeinschaft	194
6.1.1.4.8 Anmerkungen.....	195
6.1.1.5 Validitätsbedrohungen und systematische Verzerrungen	195
6.1.1.5.1 Reihenfolgeeffekte.....	195
6.1.1.5.2 Formale Antwortstile und Antworttendenzen.....	195
6.1.1.6 Auswertungskonzept	195
6.2 Datenauswertung	197
6.2.1 Itemanalyse (Objektiv-Skala).....	197
6.2.1.1 Itemschwierigkeit und Itemtrennschärfe (Objektiv-Skala).....	197
6.2.1.1.1 Schwierigkeitsanalyse	200
6.2.1.1.1.1 Items mit mittlerer Schwierigkeit und überdurchschnittlicher Streuung	202
6.2.1.1.1.2 Items mit mittlerer Schwierigkeit und unterdurchschnittlicher Streuung.....	206
6.2.1.1.1.3 Items mit extrem niedriger Schwierigkeit und variabler Streuung	209
6.2.1.1.1.4 Items mit extrem hoher Schwierigkeit und variabler Streuung.....	211
6.2.1.1.1.5 Zusammenfassung der Schwierigkeitsanalyse	212
6.2.1.1.2 Trennschärfeanalyse	213
6.2.1.1.2.1 Items mit extrem niedriger Trennschärfe	214
6.2.1.1.2.2 Items mit niedriger, signifikanter Trennschärfe	215
6.2.1.1.2.3 Items mit mittlerer, hochsignifikanter Trennschärfe	217
6.2.1.1.3 Itemselektion	220
6.2.2 Reliabilitätsanalyse (Objektiv-Skala)	225
6.2.2.1 Innere Konsistenz	227
6.2.2.2 Split-Half-Reliabilität.....	232
6.2.3 Prüfung auf Eindimensionalität (Objektiv-Skala)	234
6.2.3.1 Stichprobeneignung bzw. Opportunität der faktorenanalytischen Untersuchung.....	235
6.2.3.2 Extraktions- bzw. Abbruchkriterien	237
6.2.3.3 Einfaktorielle Lösung	242
6.2.3.4 Multifaktorielle Lösung.....	244
6.2.3.4.1 Prüfung auf generellen übergeordneten Faktor	247
6.2.3.4.1.1 Extraktions- bzw. Abbruchkriterien.....	249
6.2.3.4.1.2 Einfaktorielle Lösung	252
6.2.3.4.1.3 Zusammenfassung.....	253
6.2.4 Itemanalyse (Objektiv-Kurzskala).....	253

6.2.5 Reliabilitätsanalyse (Objektiv-Kurzskala)	254
6.2.5.1 Innere Konsistenz	254
6.2.5.2 Split-Half-Reliabilität.....	256
6.2.6 Prüfung auf Eindimensionalität (Objektiv-Kurzskala)	258
6.2.6.1 Stichprobeneignung bzw. Opportunität der faktorenanalytischen Untersuchung.....	258
6.2.6.2 Extraktions- bzw. Abbruchkriterien	260
6.2.6.3 Einfaktorielle Lösung	263
6.2.6.4 Multifaktorielle Lösung.....	265
6.2.7 Zusammenfassung.....	267
7. Hauptstudie.....	268
7.1 Methodik.....	268
7.1.1 Forschungsdesign	268
7.1.1.1 Erhebungskonzept	268
7.1.1.2 Stichprobe.....	274
7.1.1.3 Feldzugang.....	274
7.1.1.4 Operationalisierung	275
7.1.1.4.1 Glück und subjektives Wohlbefinden/Zufriedenheit	276
7.1.1.4.2 Objektiv-(Kurz)Skala	276
7.1.1.4.3 Kognitiver Pretest.....	277
7.1.1.4.4 Radiusproblem	278
7.1.1.4.5 Soziale Erwünschtheit	278
7.1.1.4.6 Korrelate Generalisierten Vertrauens.....	279
7.1.1.4.7 Standard-Skala	280
7.1.1.4.8 Demografische Standards	281
7.1.1.5 Auswertungskonzept	281
7.1.1.5.1 Kontrolltechniken.....	283
7.1.1.6 Validitätsbedrohungen und systematische Verzerrungen	284
7.1.1.6.1 Item-Nonresponse	285
7.1.1.6.2 Fragereiheneffekte.....	285
7.1.1.6.3 Antworttendenzen und Urteilsfehler.....	287
7.1.1.6.4 Verfälschbarkeit	288
7.1.1.6.5 Hawthorne-Effekt.....	290
7.1.1.7 Alternatives Forschungsdesign	290
7.2 Datenauswertung	291
7.2.1 Deskription	292

7.2.1.1 Aktuelle und allgemeine Lebenszufriedenheit	292
7.2.2 Aggregation	293
7.2.2.1 Vertrauensskalen	293
7.2.2.2 Korrelate Generalisierten Vertrauens	295
7.2.3 Validitätsanalyse	297
7.2.3.1 Inhaltsvalidität	298
7.2.3.2 Augenscheinvalidität (Face-Validity)	298
7.2.3.3 Konstruktvalidität	302
7.2.3.4 Kriteriumsvalidität	304
7.2.3.4.1 Gruppe 1: Objektiv/Standard	304
7.2.3.4.2 Gruppe 2: Standard/Objektiv	305
7.2.3.4.3 Gruppe 3: Objektiv	306
7.2.3.4.4 Gruppe 4: Standard	306
7.2.3.4.5 Zusammenfassung	309
7.2.4 Radiusproblem	309
7.2.4.1 Gruppe 1: Objektiv/Standard	310
7.2.4.2 Gruppe 2: Standard/Objektiv	311
7.2.4.3 Gruppe 3: Objektiv	313
7.2.4.4 Gruppe 4: Standard	314
7.2.4.5 Zusammenfassung	314
7.2.5 Prüfung auf Normalverteilung	315
7.2.6 Unterschiedshypothesen und Mittelwertvergleiche	317
7.2.6.1 Exkurs: Parametrische Verfahren	317
7.2.6.2 Exkurs: Nonparametrische Verfahren	318
7.2.6.3 Kontrolle von Effekten der experimentellen Situation	318
7.2.6.4 Parametrische Signifikanzprüfung	319
7.2.6.4.1 Objektiv-Skalen: PR_1 vs. PT_2 vs. PT_3	319
7.2.6.4.2 Objektiv-Kurzskalen: $PR_{1(KS)}$ vs. $PT_{2(KS)}$ vs. $PT_{3(KS)}$	321
7.2.6.4.3 Standard-Skalen: PT_1 vs. PR_2 vs. PT_4	323
7.2.6.5 Nonparametrische Signifikanzprüfung	325
7.2.6.6 Zusammenfassung	327
7.2.6.7 Effekt des Treatments (Gesamt N)	328
7.2.6.7.1 <i>Within-subject-Design</i>	329
7.2.6.7.1.1 Gruppe 1: Objektiv/Standard (PR_1 vs. PT_1)	329
7.2.6.7.1.1.1 Parametrische Signifikanzprüfung	329
7.2.6.7.1.1.2 Nonparametrische Signifikanzprüfung	332

7.2.6.7.1.2 Gruppe 2: Standard/Objektiv (PR ₂ vs. PT ₂)	336
7.2.6.7.1.2.1 Parametrische Signifikanzprüfung	337
7.2.6.7.1.2.2 Nonparametrische Signifikanzprüfung	338
7.2.6.7.1.3 Zusammenfassung	339
7.2.6.7.2 <i>Between-subject-Design</i>	340
7.2.6.7.2.1 Gruppe 3: Objektiv (PT ₃) vs. Gruppe 4: Standard (PT ₄)	341
7.2.6.7.2.1.1 Parametrische Signifikanzprüfung	341
7.2.6.7.2.1.2 Nonparametrische Signifikanzprüfung	345
7.2.6.7.2.2 Gruppenübergreifende Kontrastierung (PR ₁ vs. PR ₂ , PR ₁ vs. PT ₄ und PR ₂ vs. PT ₃)	348
7.2.6.7.2.2.1 Parametrische Signifikanzprüfung	348
7.2.6.7.2.2.2 Nonparametrische Signifikanzprüfung	351
7.2.6.7.2.3 Zusammenfassung	353
7.2.6.8 Effekt des Treatments (Identifikation der Messintention)	354
7.2.6.8.1 <i>Within-Subject-Design</i>	356
7.2.6.8.1.1 Parametrische Signifikanzprüfung	356
7.2.6.8.1.2 Nonparametrische Signifikanzprüfung	359
7.2.6.8.2 <i>Between-subject-Design</i>	362
7.2.6.8.2.1 Parametrische Signifikanzprüfung	362
7.2.6.8.2.2 Nonparametrische Signifikanzprüfung	366
7.2.6.8.3 Zusammenfassung	370
7.2.6.9 Effekt des Treatments (Identifikation der Messintention und Referenzgruppe)	371
7.2.6.9.1 <i>Within-Subject-Design</i>	372
7.2.6.9.2 <i>Between-subject-Design</i>	375
7.2.6.9.2.1 Parametrische Signifikanzprüfung	375
7.2.6.9.2.2 Nonparametrische Signifikanzprüfung	379
7.2.6.9.3 Zusammenfassung	382
7.2.6.10 Fazit der Signifikanzprüfungen	384
7.2.6.10.1 Effekte der experimentellen Situation	384
7.2.6.10.2 Auswirkungen des Treatment	386
7.2.7 Einfluss sozialer Erwünschtheit	389
7.2.7.1 Gruppe 1: Objektiv/Standard	391
7.2.7.1.1 Differenzierung nach Identifikation der Messintention	394
7.2.7.1.2 Differenzierung nach Identifikation der Messintention und Referenzgruppe	399
7.2.7.2 Gruppe 2: Standard/Objektiv	404
7.2.7.2.1 Differenzierung nach Identifikation der Messintention	407
7.2.7.2.2 Differenzierung nach Identifikation der Messintention und Referenzgruppe	411

7.2.6.3 Gruppe 3: Objektiv.....	416
7.2.6.3.1 Differenzierung nach Identifikation der Messintention	418
7.2.7.3.2 Differenzierung nach Identifikation der Messintention und Referenzgruppe	419
7.2.7.4 Gruppe 4: Standard	421
7.2.7.4.1 Differenzierung nach Referenzgruppe.....	423
7.2.7.5 Wechselseitige Kompensation von SDE- und IM-Komponente	424
8. Fazit	425
8.1 Reliabilität und Homogenität.....	425
8.2 Validität der Objektiv-Skalen	429
8.3 Messwertunterschiede	435
8.4 Einfluss reaktiver Effekte	441
8.4.1 Kontrolle von Augenscheinvalidität und Vertrauensreferenz.....	443
8.4.2 Zusammenhang von Messwertdifferenz und sozialer Erwünschtheit.....	449
8.5 Radiusproblem	457
8.6 Schlussfolgerung und Ausblick.....	458
Literaturverzeichnis	466
Anhang	493
Anhang I: Validität	494
Interkorrelation der gemittelten Skalensummenwerte von Objektiv- und Standard-Skala	494
Anhang II: Exkurse Signifikanzprüfungen	495
II.a Exkurs: Prüfung auf Normalverteilung.....	495
II.a1 Exkurs: Kolmogorov-Smirnov-Anpassungstest.....	495
II.b Exkurs: Parametrische Verfahren	496
II.b1 Exkurs: T-Test für abhängige Stichproben	496
II.b2 Exkurs: T-Test für unabhängige Stichproben	496
II.b3 Exkurs: Einfaktorielle Varianzanalyse (ANOVA)	497
II.c Exkurs: Nonparametrische Verfahren.....	499
II.c1 Test für abhängige Stichproben.....	499
II.c1.1 Exkurs: Vorzeichen-Rang-Test von WILCOXON.....	499
II.c2 Test für unabhängige Stichproben	500
II.c2.1 Exkurs: Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test.....	500
Anhang III: Partikulares vs. Generalisiertes Vertrauen	503
III.a Innerhalb der Experimentalgruppen und Messintention identifiziert (Objektiv-Skalen)	503

III.b Innerhalb Experimentalgruppen und Messintention nicht identifiziert (Objektiv-Skalen)	505
III.c Innerhalb Experimentalgruppen (Standard-Skala).....	506
III.d Zwischen den Experimentalgruppen (Objektiv-Skalen und Standard-Skala)	507
Anhang IV: Fragebogen	507
IV.a Pilotstudie	515
IV.b1 Hauptstudie (<i>Experimentalgruppe 1: Objektiv/Standard</i>)	520
IV.b2 Hauptstudie (<i>Experimentalgruppe 1: Objektiv/Standard</i>)	521
IV.c1 Hauptstudie (<i>Experimentalgruppe 2: Standard/Objektiv</i>).....	524
IV.c2 Hauptstudie (<i>Experimentalgruppe 2: Standard/Objektiv</i>).....	526
IV.d Hauptstudie (<i>Experimentalgruppe 3: Objektiv</i>).....	531
IV.e Hauptstudie (<i>Experimentalgruppe 4: Standard</i>).....	534
Anhang V: Häufigkeitsverteilungen (Pilotstudie)	534
V.a Rohskala „Generalisiertes Vertrauen“	534

1. Einleitung

Aus den *Naturwissenschaften*:

„Beim Bohrschen Atommodell wurde das Elektron als ein bewegtes Teilchen angesehen. Um seine Bahn zu berechnen, muss man zu einem gegebenen Zeitpunkt gleichzeitig seine Geschwindigkeit und seinen Aufenthaltsort kennen. Nach der **Unschärferelation** von Heisenberg (1926) gilt jedoch:

Es ist grundsätzlich unmöglich, von einem Objekt gleichzeitig den genauen *Aufenthaltsort* und den *Impuls* zu bestimmen.

Die Lage von Körpern sehen wir mit Hilfe von Licht. Um ein so kleines Objekt wie ein Elektron zu orten, ist Licht mit sehr kurzer Wellenlänge notwendig. Kurzwelliges Licht hat eine hohe Frequenz und ist sehr energiereich. Wenn es das Elektron trifft, erteilt es ihm einen zusätzlichen Impuls. Der Versuch das Elektron zu orten, verändert seinen Impuls drastisch. Energieärmere Photonen würden den Impuls weniger beeinflussen, wegen der zugehörigen größeren Wellenlänge könnte man damit das Elektron jedoch nur ungenau orten. Nach Heisenberg ist die Unschärfe (Ungenauigkeit) bei der Bestimmung des Ortes, Δx , mit der Unschärfe des Impulses, $\Delta(mv)$, verknüpft durch:

$$\Delta x * \Delta(mv) \geq \frac{h}{4\pi}$$

Für gewöhnliche Objekte ist wegen der relativ hohen Masse m die Unschärfe der Messung ohne Bedeutung, bei kleinen Teilchen wie Elektronen jedoch so erheblich, dass Aussagen über Elektronenbahnen in Atomen hoffnungslos sind.“ (Mortimer/Müller (2001), S. 70f)

Aus den *Sozialwissenschaften*:

„Legt man einer Person einen Fragebogen vor, mittels dessen sie sich selbst, ihre Einstellungen oder ihre Selbstkonzepte beschreiben soll, so wird die betreffende Person aufgefordert, ihr subjektives Urteil zu etwas im Grunde sehr ‚Persönlichem‘ abzugeben. Damit liegt auf der Hand, dass Messungen mittels Persönlichkeitsfragebogen niemals objektiv sind, sondern bei näherem Hinsehen in mehrfacher Weise subjektiv: Einmal geht es um einen subjektiven Gegenstand, also um etwas, über das zumeist nur das Individuum selbst Bescheid wissen kann, und zum anderen legt die antwortende Person bei der Beurteilung dieses Gegenstandes einen subjektiven Maßstab an. Erfährt die urteilende Person somit, dass ihr subjektives, prinzipiell offensichtlich nicht nachprüfbares Urteil gefragt ist, und weiß sie beispielsweise, dass sie beobachtet wird und Objekt einer wissenschaftlichen Untersuchung ist, so steht ihr die Möglichkeit offen, ihre Reaktionen bzw. Antworten auf den Fragebogen absichtlich oder unabsichtlich in bestimmter Richtung zu beeinflussen.“ (Mummendey/Grau (2014), S. 165)

Aus dem *täglichen Leben*:

„Wir kennen das Phänomen aus dem Alltag: Jemand hält eine Fotokamera auf uns, um uns zu fotografieren. Automatisch setzen wir ein Lächeln auf, um vorteilhaft im Bild zu erscheinen. Alternativ kann man sich auch abwenden, um die Aufnahme zu verhindern, wenn einem nicht danach ist fotografiert zu werden. (...) Die beobachteten Personen wissen, dass sie beobachtet werden und verhalten sich (möglicherweise) anders, als sie sich verhalten würden, wenn sie sich nicht beobachtet fühlten. (...) Wer lächelt, wenn er fotografiert wird, obwohl ihm gerade nicht danach zumute ist, gibt eine falsche Auskunft über die eigene emotionale Befindlichkeit. Wer sich abwendet, verhindert, dass man überhaupt etwas über sie erfährt, außer dass sie nicht beobachtet oder befragt werden will. Auf diese Weise entsteht eine eigene (künstliche) Wirklichkeit durch den Einsatz von Methoden und Instrumenten (hier: Fotokamera), die es ohne deren Einsatz (hier: das Fotografieren) nicht gegeben hätte und die nicht mit der sozialen Wirklichkeit übereinstimmt, welche auf alltägliche Weise konstruiert wird.“ (Scholl (2013), S. 79f)

Auf den ersten Blick scheinen die drei hier beschriebenen Begebenheiten zunächst keine Gemeinsamkeit aufzuweisen. Während zumindest die Sozialwissenschaften und die Alltagsepisode eine gewisse Ähnlichkeit aufweisen, scheint es sich insbesondere bei der

1. Einleitung

Heisenberg'schen Unschärferelation um eine vergleichsweise abstrakte Problematik zu handeln. Tatsächlich, wenngleich auch auf unterschiedlichen Abstraktionsebenen, beschreiben allerdings alle drei Zitate das gleiche Phänomen: *Reaktivität. Das Forschungsobjekt reagiert auf die Art und Weise der Messung, woraufhin das Messergebnis verfälscht wird.* Messwertverzerrungen bzw. die Unschärfe von Messungen stellen insofern ein universales Problem dar.

Im Rahmen der Physik manifestiert sich Reaktivität in der Tatsache, dass es unmöglich ist, „Ort und Impuls eines Teilchens gleichzeitig mit beliebiger Genauigkeit vorherzusagen“ (Meschede (2002), S. 567). Nach der Heisenberg'schen Unschärferelation „reagiert“ ein Teilchen beispielsweise auf den Versuch seinen Ort zu bestimmen, indem es seinen Impuls verändert. Aus der Messung seines Impulses resultiert dagegen eine Unschärfe des Ortes, an dem sich das Teilchen im Moment des Messens befunden hat.

Innerhalb der empirischen Sozialforschung können reaktive Effekte verschiedene Formen annehmen und lassen sich im Wesentlichen in Bedrohungen für die interne und die externe Validität wissenschaftlicher Studien unterteilen. Während interne Validität gegeben ist, wenn tatsächlich die unabhängige Variable (Treatment bzw. UV) für die Variation der abhängigen Variable (AV) verantwortlich ist, beschreibt externe Validität die Möglichkeit der Generalisierung bzw. Verallgemeinerung von Resultaten einer Studie. Externe Validität ist somit dann gegeben, wenn sich die Ergebnisse einer empirischen Untersuchung auf andere Personen in anderen Kontexten übertragen lassen oder die im Rahmen künstlicher Laborsituationen gefundenen Forschungsergebnisse auch in der realen Welt Gültigkeit besitzen.

Reaktive Effekte können als Folge der experimentellen Situation oder des Messens auftreten. In ersterem Fall werden sie im Wesentlichen durch die Künstlichkeit der Untersuchungssituation oder das Forschungsdesign der jeweiligen Studie hervorgerufen. Beispiele hierfür sind sogenannte *History-* oder Reifungsprozesse, die Ereignisse und intrapersonale Prozesse beschreiben, die zusätzlich zur UV Einfluss auf die AV ausüben. Daneben existieren Messeffekte, wie Positions- bzw. Reihenfolgeeffekte oder sogenannte *Carry-over*-Effekte. Diese beschreiben Veränderungen der AV, die auf die Auswirkungen eines ersten Messvorgangs und nicht auf den Einfluss des Treatments zurückzuführen sind. Der *Hawthorne*-Effekt beschreibt Messwertverzerrungen, die aus dem Bewusstsein der Probanden resultieren, Gegenstand einer wissenschaftlichen Untersuchung zu sein. *Testing and Instrumentation* beschreibt Messfehler, die sich infolge von Lerneffekten bzw. Erfahrung der Probanden mit dem Messinstrument oder der Inkonsistenz des Messinstruments über verschiedene Messzeitpunkte und der daraus resultierenden mangelnden Vergleichbarkeit der Messergebnisse, ergeben. *Diffusion of Treatments* meint, dass Versuchs- und Kontrollgruppe gleichermaßen durch das Treatment beeinflusst werden und so eine Isolierung des Treatmenteffektes auf die AV verhindert wird. Schließlich können auch Stichproben- bzw. Selektionsfehler und systematische Ausfälle von Probanden, die sogenannte *Mortality*, für Messwertverfälschungen verantwortlich sein. So können beispielsweise bestimmte Bevölkerungsgruppen innerhalb der zugrunde gelegten Stichprobe über- oder unterrepräsentiert sein oder bestimmte Probanden nehmen nur an einzelnen Teilen der empirischen Untersuchung teil.

Reaktive Effekte des Messens beschreiben in der Regel sogenannte Antwortverzerrungen oder *Response errors*, die sich hauptsächlich aus der Reaktion der Probanden auf die Frage- und Antwortvorgaben bzw. Itemformulierungen eines Messinstrumentes, der Itemreihenfolge innerhalb eines Messinstrumentes oder der Interaktion von Proband und Interviewer/Versuchsleiter ergeben. Fragereiheneffekte ergeben sich aus der relativen Position

1. Einleitung

einzelner Items zu anderen Items. Je nachdem an welcher Stelle im Untersuchungsinstrument bestimmte Themen behandelt werden oder in welcher Reihenfolge Themen im Rahmen des Untersuchungsinstrumentes abgefragt werden, können sich Unterschiede im Antwortverhalten ergeben. Prominente Beispiele hierfür sind der *Halo*-Effekt oder sogenannte *Primacy*- und *Recency*-Effekte. Versuchsleitereffekte äußern sich dergestalt, dass die Probanden versuchen zu antizipieren, was der Versuchsleiter von ihnen erwartet bzw. was die Erwartungen des Versuchsleiters bezüglich der Ergebnisse der Studie sein mögen und passen ihr Antwortverhalten entsprechend an. Die Folge kann gleichermaßen eine überwältigende Bestätigung wie ausdrückliche Widerlegung der Erwartungen des Versuchsleiters sein. Beides findet allerdings keine Entsprechung in den tatsächlichen realen Verhältnissen. Der *Pygmalion*-Effekt beschreibt möglicherweise unbewusste jedoch aktive Versuche des Versuchsleiters, das Antwortverhalten der Probanden in Richtung seines Wunschergebnisses zu beeinflussen. Antworttendenzen wie *Akquieszenz* (Zustimmungstendenz) lassen die Probanden unabhängig vom Inhalt der Frage eine Zustimmung formulieren. Bewusste Verfälschungen des Antwortverhaltens stellen sich als sogenanntes *fake good* oder *fake bad* dar. Hierbei versuchen sich die Probanden entweder möglichst positiv oder möglichst negativ bezüglich der jeweiligen Zieldimension darzustellen. Zu dieser Kategorie reaktiver Effekte zählt u.a. die Tendenz sozialerwünschte Antworten zu geben. Messwertverzerrungen können sich aber auch als *Item-Nonresponse*, d.h. der Verweigerung einer Antwort, *Meinungslosigkeit*, d.h. der Angabe von „Weiß-nicht“ oder sogenannte *Non-Attitudes*, d.h. der Abgabe einer Wertung, obwohl der Proband sich keine Meinung zum Forschungsgegenstand gebildet hat, manifestieren.

Die hier vollzogene Auflistung potentieller reaktiver Effekte ist keineswegs erschöpfend, vielmehr existiert noch eine Vielzahl weiterer Quellen für Messwertverfälschungen. Reaktive Effekte, die aus der experimentellen Situation herrühren, sind dabei eher als Bedrohungen der internen Validität zu verorten, während reaktive Effekte, die auf den Messvorgang zurückzuführen sind, eher Bedrohungen der externen Validität darstellen. Jedoch kann der Übergang von reaktiven Effekten, die die interne bzw. externe Validität betreffen, als fließend angesehen werden. Manche Fehlerquellen wirken sich sogar gleichermaßen auf die interne wie die externe Validität aus.

Entsprechend der Mannigfaltigkeit reaktiver Effekte existiert in der empirischen Sozialforschung eine Fülle von Verfahren zur Vermeidung, Reduktion oder Eindämmung reaktiver Effekte und ihrer Konsequenzen. Klassische Techniken zur Kontrolle von Störfaktoren sind die Elimination oder das Konstanthalten potentieller Störgrößen sowie die Bildung von Kontrollgruppen per Parallelisierung oder randomisierter Zuteilung der Probanden. Solche Maßnahmen erweisen sich allerdings insbesondere in Bezug auf Messwertverzerrungen, die dem Antwortverhalten der Probanden entspringen, als unwirksam.

Hierbei ist jedoch festzuhalten, dass nur eine besondere Gruppe von Erhebungsverfahren, sogenannte direkte Messverfahren, überhaupt zur Anfälligkeit für reaktive Effekte neigt. Messmethoden dieser Verfahrensgruppe werden auch als explizite oder reaktive Messverfahren bezeichnet und machen eine bewusste Aufmerksamkeit der Probanden für den Forschungsgegenstand und die Art und Weise seiner Erfassung notwendig. Dabei enthüllen sie ausdrücklich ihre Messintention und setzen als Selbstbeurteilungs- bzw. Selbstauskunftsmaße eine explizite Artikulation der Einstellungen, Urteile und Meinungen bzw. eine Selbstoffenbarung seitens der Probanden voraus. Es ist diese Transparenz des Messprozesses, die direkte Messverfahren für reaktive Effekte prädestiniert.

1. Einleitung

Als Alternative oder Pendant zu direkten Messverfahren wurden daher sogenannte indirekte Messverfahren entwickelt, die keine bewusste Aufmerksamkeit der Probanden für den Messprozess erfordern. Vielmehr wird versucht, reaktive Effekte zu verhindern, indem man Selbsteinschätzungen und unmittelbare verbale Statements bezüglich der interessierenden Zieldimension vermeidet, den Kontakt zwischen Proband und Forscher auf ein Minimum reduziert, Datenerhebungen im Rahmen der Lebenswelt der Probanden vornimmt und die jeweilige Messintention möglichst undurchsichtig gestaltet. Indirekte Messverfahren differenzieren sich in eine Reihe von Subklassen auf, die nicht immer alle der hier genannten Kriterien erfüllen. Gemeinsam ist jedoch allen, dass sie eine vergleichsweise geringe Anfälligkeit für gezielte Verfälschungen ihrer Messwerte aufweisen. Physiologische Messungen erfassen die Lebensvorgänge innerhalb des menschlichen Organismus, um Rückschlüsse auf die psychische Verfassung des jeweiligen Probanden zu ziehen. Nicht-reaktive Verfahren versuchen zu verhindern, dass den Probanden die Einbindung in einen Messprozess bewusst wird, indem sie den direkten Kontakt zum Forscher ausschalten und Datenerhebungen in der vertrauten Alltagsumgebung der Probanden stattfinden lassen. Implizite Maße setzen keine bewusste Aufmerksamkeit für den Forschungsgegenstand voraus und erfassen die Einstellungen der Probanden, ohne diese direkt um eine verbale Äußerung ihrer Einstellungen zu bitten. Projektive Verfahren verwenden vergleichsweise unstrukturierte und mehrdeutige Testvorlagen, die von den Probanden interpretiert, rezipiert, vervollständigt oder angeordnet werden sollen, um aus den Äußerungen der Probanden Rückschlüsse auf deren Persönlichkeitsstruktur und psychische Verfasstheit ziehen zu können. Objektive Tests stellen Leistungsmessungen, Simulationen des täglichen Lebens oder Fragebogen dar, die das Verhalten der Probanden in einer standardisierten Situation unter Rückgriff auf Items mit geringer Augenscheinvalidität bezüglich der Zieldimension, und ohne dass die Probanden eine Selbstbeurteilung vollziehen müssen, erfassen.

Abgesehen von ihrer Anfälligkeit für reaktive Effekte weisen direkte Messverfahren eine Vielzahl von Vorzügen auf, die der Grund für ihre weitverbreitete Nutzung sind. So erlauben Selbstauskunftsmaße eine einfache Adaption auf viele verschiedene Inhalte, ermöglichen eine spezifische Abfrage der interessierenden Zieldimension und entsprechende Untersuchungsinstrumente lassen sich vergleichsweise schnell erstellen sowie zeitnah in der Praxis einsetzen. Zusammengefasst weisen direkte Messverfahren eine besondere Flexibilität und Benutzerfreundlichkeit auf, allerdings zum Preis potentieller Messwertverfälschungen. Indirekte Messverfahren sind dagegen weitestgehend immun gegenüber reaktiven Effekten und dazu in der Lage, auch Facetten eines Forschungsgegenstandes abzubilden, zu denen sich mittels direkter Messverfahren kein Zugang eröffnet. Jedoch sind die Messkonzepte und -prinzipien indirekter Messverfahren oft auf spezifische Phänomene limitiert. Zusätzlich gestaltet sich deren Operationalisierung in vielen Fällen recht aufwändig. Indirekte Messverfahren bieten somit Verfälschungssicherheit, das Potential für einen erweiterten Erkenntnisgewinn und eine differenziertere Abbildung des interessierenden Konstruktes, jedoch auf Kosten von Vielseitigkeit und Anwenderfreundlichkeit.

Direkte Messverfahren scheinen somit über Eigenschaften zu verfügen, die indirekten Messverfahren fehlen und umgekehrt.

Es ist anzunehmen, dass aus der **Integration beider Verfahrensgruppen** eine Erhebungsmethode resultieren sollte, die die Vorzüge beider Verfahrensklassen vereinigt und

1. Einleitung

auf diese Weise ihre jeweiligen Defizite ausgleicht. Entsprechend müsste ein hybrides Messverfahren ähnlich flexibel und intuitiv einsetzbar sein wie direkte Messverfahren, während es gleichzeitig kaum Messwertverfälschungen infolge reaktiver Effekte aufweist und eine besonders differenzierte Messung der interessierenden Zieldimension ermöglicht.

Dieses explorative Forschungsprojekt soll nun als erstes das Ziel verfolgen, einen solchen Hybriden aus direkten und indirekten Messverfahren zu entwickeln, indem es beide Verfahrensklassen miteinander verbindet. Als Ansatzpunkt für dieses Unternehmen wurde die Fragebogenvariante Objektiver Tests ausgewählt, da diese ein indirektes Messverfahren darstellt, dessen Messprinzip ein hohes Maß an Übereinstimmung mit dem Messkonzept von direkten Messverfahren besitzt. Entsprechend sollte eine Vereinigung direkter und indirekter Messverfahren anhand dieser Erhebungsmethode vergleichsweise einfach gelingen. Daneben werden in empirischen Studien mit weitem Abstand am häufigsten verschiedene Formen des Interviews für die Datenerhebung eingesetzt. Es ist daher nur folgerichtig, auch ein hybrides Messinstrument als Fragebogen zu konzipieren, um der großen Beliebtheit und den breiten Einsatzmöglichkeiten des Survey-Designs Rechnung zu tragen. Auf einer rein technischen Ebene bedeutet die Auswahl des Fragebogendesigns eine Sicherstellung der Objektivität der Messung sowie eine Vereinfachung der Bestimmung von Reliabilität und Validität des sich ergebenden Messinstrumentes.

Das zweite Ziel dieses Forschungsprojektes besteht darin zu überprüfen, ob durch die Verwendung des Hybriden aus direkten und indirekten Messverfahren tatsächlich die prognostizierte Verbesserung der Messgüte gegenüber direkten Messverfahren bewirkt werden kann. Mit anderen Worten: Sind die Messwerte des hybriden Messinstrumentes weniger stark durch reaktive Effekte verzerrt als die Messwerte eines traditionellen Fragebogens, der die gleiche Zieldimension abbildet?

Die empirische Untersuchung gliedert sich dabei in zwei Phasen. So wird im Rahmen der Pilotstudie das hybride Messinstrument, welches nachfolgend auch als Objektiv-Skala bezeichnet wird, entlang einer Querschnittsbefragung konstruiert. Als Platzhaltervariable bzw. inhaltliches Vehikel wurde dabei das theoretische Konstrukt „*generalisiertes Vertrauen*“ verwendet. Besonderer Vorteil dieser Zieldimension ist es, dass mit der *Trust in People-Scale* ein Standardinstrument zur Erfassung *generalisierten Vertrauens* existiert, welches für die spätere Kontrastierung von Hybride und direktem Messverfahren herangezogen werden kann. Zudem haben empirische Untersuchungen im Rahmen der Vertrauensforschung gezeigt, dass die Messung *generalisierten Vertrauens* mit Hilfe des Standardinstrumentes nicht frei vom Einfluss reaktiver Effekte ist. Insofern handelt es sich bei *generalisiertem Vertrauen* um eine Zieldimension, deren Messwerte verfälscht werden können.

In der zweiten Phase, der Hauptstudie, wird anhand eines Survey-Experimentes untersucht, ob sich bei der Erfassung *generalisierten Vertrauens* Messwertunterschiede zwischen der Objektiv-Skala und der Standard-Skala (*Trust in People-Scale*) ergeben und inwiefern diese auf reaktive Effekte zurückzuführen sind. Hierbei steht die Neigung der Probanden zu sozialerwünschtem Antwortverhalten im Fokus des Interesses. In Bezug auf *generalisiertes Vertrauen* manifestiert sich dieses als *fake good* bzw. die Tendenz, sich möglichst vertrauensvoll darzustellen. Mit anderen Worten: Es gilt als sozialwünschenswert, ein besonders hohes Vertrauensniveau zu erreichen und nicht als misstrauisch zu gelten.

Zusammengefasst verfolgt die Pilotstudie somit im Wesentlichen die folgende Fragestellung:

1. Einleitung

„Lässt sich ein Messinstrument im Fragebogendesign entwickeln, das eine Verbindung aus direkten und indirekten Messverfahren darstellt und wie muss dieses beschaffen sein?“

Während sich die Hauptstudie vornehmlich mit dieser Fragestellung befasst:

„Lässt sich mit Hilfe des Hybriden aus direkten und indirekten Messverfahren eine Verbesserung der Messgüte im Kontrast zu direkten Messverfahren bewirken?“

Die vorliegende Arbeit erbringt dabei mehrere Forschungsleistungen. So macht die besondere Architektur der Objektiv-Skala eine detaillierte theoretische Aufarbeitung *generalisierten Vertrauens* notwendig, woraus sich die Formulierung einer integrativen Perspektive auf das Vertrauensphänomen ergibt. Daneben erfolgt die Konstruktion eines neuen Erhebungsinstrumentes zur Erfassung *generalisierten Vertrauens* auf der Basis einer Integration von direkten und indirekten Messverfahren. Zuletzt wird der Hybride aus direkten und indirekten Messverfahren mit einem bestehenden direkten Messverfahren kontrastiert, um festzustellen, ob und in welchem Ausmaß aus der Verwendung der Objektiv-Skala eine Reduktion von Messwertverzerrungen gegenüber der Standard-Skala resultiert bzw. sich insgesamt eine Verbesserung der Messgüte einstellt. Zusätzlich wird untersucht, ob das hybride Messinstrument einen Beitrag zur Lösung des sogenannten Radiusproblems der Vertrauensforschung leisten kann. Im Einzelnen bedeutet das, ob im Rahmen der Vertrauensmessung mit Hilfe der Objektiv-Skala eine nachhaltige inhaltliche Trennung zwischen den voneinander zu unterscheidenden Zieldimensionen *partikulares* und *generalisiertes Vertrauen* bewerkstelligt werden kann.

2. Theorie: Methodik

2.1 Reaktivität

Das Konzept der „Reaktivität“ geht zurück auf CAMPBELL und fasst eine ganze Reihe von Verzerrungseffekten zusammen. (Bungard/Lück (1974), S. 88ff; Hartmann (1991), S. 22ff) Nach CAMPBELL manifestiert sich Reaktivität dabei folgendermaßen:

„(...) the respondents can become aware that they are participating in an experiment, and this awareness can have an interactive effect, in creating reactions to X which would not occur had X been encountered without this „I’m a guinea pig“ attitude.“ (Campbell (1957), S. 308)

Mit anderen Worten: Reaktivität beschreibt im Allgemeinen die Tatsache, dass Personen im Rahmen von Erhebungssituationen auf die jeweiligen Begleitumstände und die durch den Versuchsleiter bzw. Interviewer oder das Untersuchungsinstrument gesetzten Stimuli reagieren. (Schnell/Hill/Esler (2008), S. 353) Begreift man die Erhebungssituation ganz allgemein als sozialen Interaktionsprozess, so wird deutlich, dass Reaktivität keineswegs nur ein Phänomen (sozial-)wissenschaftlicher Methodologie darstellt, sondern sich auch im Alltag beobachten lässt. (Scholl (2013), S. 79f)

„Soziales Handeln aber soll ein solches Handeln heißen, welches seinem von dem oder den Handelnden gemeinten Sinn nach auf das Verhalten anderer bezogen wird und daran in seinem Ablauf orientiert ist.“ (Weber (1980), S. 1)

Reaktivität ist in diesem Sinne integraler Bestandteil sozialen Handelns, insofern diesem ein interaktives Moment innewohnt. So bedeutet soziales Handeln immer auch eine Reaktion auf die Handlungsweisen des Gegenübers, wobei die Ausgestaltung dieser Reaktion nicht unabhängig vom Verhalten dieses Gegenübers stattfindet. Die Folge kann ein Verhalten bzw. Handeln sein, das einzig und allein zu dem Zweck ausgeführt wird, den Erwartungen des anderen zu entsprechen. Dabei existiert es mehr oder weniger losgelöst vom tatsächlichen Verhalten des ausführenden Akteurs. Diese Tendenz zur Verfälschung der eigenen Verhaltensweise ist es, was als Problem der Reaktivität beschrieben werden kann.

Vom Problem der Reaktivität spricht man, wenn das im Rahmen der sozialen Situation gezeigte Verhalten eine Abweichung vom wahren Wert bedeutet. In einem solchen Fall resultiert eine Verzerrung sozialer Wirklichkeit, was im (sozial-)wissenschaftlichen Kontext mit einer Bedrohung der Validität empirischer Ergebnisse und letztendlich einer Ungültigkeit wissenschaftlicher Erkenntnis bzw. der Erschaffung von Singularitäten oder methodischen Artefakten gleichzusetzen ist. Von besonderer Tragweite ist es, wenn durch reaktive Effekte nicht nur eine Verzerrung der Realität bewirkt wird, sondern eine vollkommen künstliche Wirklichkeit hervorgebracht wird. (Scholl (2013), S. 85)

Mit dem system- und kommunikationstheoretischen Reaktivitätsansatz von MERTEN und dem rollen- und handlungstheoretischen Reaktivitätsansatz von ESSER existieren zwei konkurrierende theoretische Erklärungen für das Phänomen der Reaktivität. (Scholl (2013), S. 82ff)

MERTEN geht zunächst davon aus, dass bei der sozialwissenschaftlichen Datenerhebung bzw. der Erfassung sozialer Wirklichkeit zwei miteinander verschränkte Kommunikationsprozesse wirksam werden. Der erste Kommunikationsprozess ist dabei mit der sozialen Wirklichkeit selbst gleichzusetzen und der zweite Kommunikationsprozess beschreibt die Vermittlung sozialer

2. Theorie: Methodik

Wirklichkeit im Rahmen der Erhebungssituation. Bei der Erhebung sozialer Daten müssen beide Kommunikationsprozesse „notwendig miteinander interferieren“ (Merten (2005), S. 109), wobei diese Interferenz von Erhebungsinhalt und Vermittlung des Inhaltes mit Reaktivität gleichzusetzen ist. Insofern gilt, „dass soziale Wirklichkeit niemals pur, damit auch niemals konstant oder fixiert als Datum erhebbar ist, sondern allenfalls als Wirklichkeit erfasst werden kann, die bereits durch eine andere Wirklichkeit (Erhebungssituation) gebrochen ist“ (Merten (2005), S. 110). Mit anderen Worten: Erhebungs- und erhobene Wirklichkeit sind analytisch nicht voneinander zu trennen. (Scholl (2013), S. 84) Dies ist damit zu begründen, dass die Erhebungssituation selbst Teil der sozialen Wirklichkeit ist bzw. im jeweiligen Moment aktuelle soziale Wirklichkeit darstellt. MERTEN schließt hieraus, dass Reaktivität notwendigerweise immer Bestandteil der Erhebung sozialer Wirklichkeit ist und durch entsprechende Maßnahmen nicht nachhaltig ausgeschaltet werden kann. (Merten (2005), S. 113)

Bedingung für Reaktivität ist weniger die Interaktivität als die zeitliche, sachliche und soziale Reflexivität der Erhebungssituation. Die Reflexivität der Zeitdimension bezieht sich darauf, dass die Gegenwart nicht unabhängig von der Vergangenheit existiert bzw. „Kommunikationsprozesse direkt und unmittelbar auf sich selbst zurückwirken“ (Merten (2005), S. 111). Form und Ergebnis von Informationsverarbeitung sind demnach eng an die im Vorhinein zur Verfügung stehenden Informationen gebunden. Die Reflexivität der Sachdimension ergibt sich aus der Relevanz von Meta-Informationen, die jede Kommunikation begleiten und deren Interpretationsrahmen festschreiben. Solche Meta-Informationen werden dabei sowohl über einen verbalen als auch einen nonverbalen Kanal vermittelt. Die Reflexivität der Sozialdimension resultiert aus der Orientierung an Anderen im Rahmen des Kommunikationsprozesses, wobei keine unmittelbare Gegenwärtigkeit dieser Anderen vorausgesetzt ist. (Merten (2005), S. 111f)

Reaktivität tritt insbesondere dann auf, wenn die Erhebungssituation Reflexivität im obengenannten Sinne erfordert oder fördert, wobei zu bedenken ist, „dass sozialwissenschaftliche Erhebungsverfahren nicht nicht reaktiv sein können“ (Merten (2005), S. 120). Reaktivität ist dabei allerdings nicht per se negativ zu bewerten, da sie eine Grundvoraussetzung für die Erfassung von sozialer Realität bildet:

„Notwendig ist Reaktivität insofern, als der Kommunikationsprozess, der zwischen Erhebendem und Erhebungsobjekt initiiert werden muss, die Vermittlung des inhaltlichen Kommunikationsprozesses leisten muss, der sowohl *reell* (Erhebung von unmittelbar ablaufendem sozialen Handeln) als auch *virtuell* (Erhebung von vorgestelltem oder reflektiertem sozialen Handeln) angelegt sein kann.“ (Merten (2005), S. 120)

Zum Problem wird Reaktivität, wenn die Erfassung sozialer Wirklichkeit durch sie verzerrt wird:

„Störend ist alle Reaktivität, die über die notwendige Anfangsphase und die Überwachung des Datentransfers vom inhaltlichen in den Vermittlungskontext durch den Erhebenden hinaus bestehen bleibt. Sie kann vom Erhebenden, vom Untersuchungsobjekt, aber auch von den Daten selbst induziert werden.“ (Merten (2005), S. 121)

Die Verzerrung der Daten durch Reaktivität ist dabei einem Mechanismus geschuldet, der bewirkt, dass die im Rahmen der Erhebungssituation getätigten inhaltlichen Aussagen mehr oder weniger bewusst selektiert werden. Geordnet nach Reflexivitätsdimension nennt MERTEN die nachfolgenden Reaktivitätsantezedenzen (Merten (2005), S. 121f):

- (1) *Zeitdimension*: Konditionierung des Verhaltens oder Veränderung von relevantem Wissen, Meinungen oder Einstellungen des Rezipienten durch den Erhebungsprozess.

2. Theorie: Methodik

- (2) *Sachdimension*: verbale oder nonverbale Bewertungen des Kommunikators; Neutralitätsverletzung durch Kommunikator oder Rezipient; Interpretationsversuche seitens des Rezipienten; Kommunikation von Erwartungen hinsichtlich richtigem oder falschem Verhalten; Normen- oder Wertänderungen, die eine Modifikation der Daten während des Erhebungsprozesses bewirken.
- (3) *Sozialdimension*: Anwesenheit bzw. Einflussnahme Dritter; Veränderungen von Ansichten des Rezipienten während des Erhebungsprozesses, die aus einer Orientierung an Anderen resultieren; Annahmen hinsichtlich des Kommunikators bzw. Orientierung am Verhalten des Kommunikators

Zu bemerken ist, dass sich Reaktivität dann am stärksten auf die Güte der Daten auswirkt, wenn sie durch eine Verknüpfung der verschiedenen Reflexivitätsdimensionen hervorgerufen wird.

(Merten (2005), S. 122):

Die Tatsache, dass „soziale Realität nicht unabhängig von der sozialen Realität des Forschungsprozesses erfasst werden kann“ (Scholl (2013), S. 85) hat in letzter Instanz auch Einfluss auf die Überprüfbarkeit von Theorien. Unter dem Einfluss von Reaktivität empirisch geprüfte Theorien müssten nämlich den gleichen Grad an Verzerrung aufweisen, wie die zugrunde gelegten Daten und stellen deshalb nur eine unzureichende Beschreibung der sozialen Wirklichkeit dar. Nach MERTEN lässt sich dieses Dilemma nur lösen, indem man auch der Theorie Reflexivität zugesteht. (Merten (2005), S. 122ff)

Nach ESSER wird als Reaktivität „in einem weiten Sinn die Reaktion einer Untersuchungsperson verstanden, welche von dem methodisch erforderlichen Verhalten – nämlich die bloße Reaktion auf den Untersuchungsstimulus – abweicht“ (Esser (1975b), S. 258). Reaktivität weist dabei graduelle Unterschiede auf, die von einer grundlegenden Bedingung für den Datenabruf bis hin zu einer systematischen und absichtsvollen Reaktivität reichen, nach der das Untersuchungsobjekt keinen neutralen Datenträger mehr darstellt. (Esser (1975a), S. 98f; Esser (1975b), S. 265; Esser (1986), S. 326). Insbesondere in letzterem Fall folgt das Antwortverhalten der Untersuchungsobjekte einem Kosten-Nutzen-Kalkül, welches sich sowohl „an den Präferenzen, Zielsetzungen und normativen Einbindungen der Personen“ (Esser (1986), S. 320) orientiert als auch dem „Hintergrund der Perzeption bzw. Kognition der Situation und den damit jeweils vorliegenden Ambiguitäten, Risiken und Möglichkeiten“ (Esser (1986), S. 320) Rechnung trägt.

Ausgangspunkt der Überlegungen von ESSER ist die Annahme, dass zur Etablierung eines Forschungskontaktes die Einwilligung der Untersuchungsobjekte in eine Interaktionsbeziehung erforderlich ist, wobei dies umso einfacher zu bewerkstelligen ist, je stärker die Motivation zum Eingehen von Interaktionen im Allgemeinen ausgeprägt ist. Die Motivation zum Eintritt in eine Interaktionsbeziehung kann dabei als Motivation zur Erfüllung einer bestimmten Rollenverpflichtung bzw. als Motivation zur Erlangung einer aus der Rollenerfüllung resultierenden Gratifikation verstanden werden. Welche wechselseitigen Verhaltenserwartungen die jeweilige Situation beinhaltet, wird anhand der Situationsstrukturierung antizipiert, wobei auch wechselseitige Missverständnisse die Folge sein können. (Esser (1975b), S. 258ff)

„(...) die Motivation zur Erfüllung bestimmter Umwelterwartungen ist gegründet auf die Aspiration zur Erlangung von Gratifikationen bzw. der Vermeidung negativer Sanktionen. Werden Personen nun konfrontiert mit Situationen, die einen hohen Grad an Erwartungsambiguität aufweisen, und ist eine Strukturierung der Situation von außen nur bruchstückhaft erfolgt, so wird die Situationsstrukturierung nur als individuelle Leistung möglich sein, und zwar so, daß das Erlangen von Gratifikationen möglichst wahrscheinlich wird.“ (Esser (1975b), S. 264)

Die Gratifikation entspricht dabei zunächst sozialer Anerkennung, die sich aus der Funktionserfüllung des Untersuchungsobjektes im Rahmen des Forschungsprozesses ergibt. Entsprechend orientiert sich die Rollenselektion des Untersuchungsobjektes an den als gewünscht vermuteten Reaktionen. (Esser (1975a), S. 86ff, S. 91ff; Esser (1975b), S. 261f) Aus der Tatsache, dass es sich bei der Situationsdefinition und den daraus resultierenden Vermutungen hinsichtlich der gültigen Rollenerwartungen lediglich um Antizipationen bzw. Annahmen handelt, folgt, dass die Rollenselektion bereits eine vom Inhalt losgelöste Reaktion auf den Forschungsprozess darstellt. Dabei weicht die Einschätzung des Untersuchungsobjektes umso mehr von einer korrekten Einstufung ab, je unbestimmter die Situation ist bzw. je weniger die Situation durch die Versuchsleitung strukturiert wird. Es gilt daher: *„Reaktivität ist der Preis der Rollenübernahme.“* (Esser (1975b), S. 265) Dies ist insofern einsichtig, als dass die Erhebung von Daten innerhalb der Forschungssituation erst durch Auskünfte der Forschungsobjekte bzw. deren Reaktion auf das Forschungsinstrument ermöglicht wird.

Gleichzeitig kann systematische intentionale Reaktivität eine Bedrohung der externen und internen Validität nach sich ziehen. In diesem Fall erweitert sich das Interesse des Forschungsobjektes von der bloßen Erlangung sozialer Anerkennung hin zur Nutzenmaximierung. (Esser (1975b), S. 265; Esser (1986), S. 319) Reaktivität ist dann „die Folge von Konsequenzenbefürchtungen der partikularistischen Art“ (Esser (1975b), S. 266), wobei Konsequenzen gleichermaßen Sanktionen wie Gratifikationen einschließen. So setzen Forschungskontakte bzw. sozialwissenschaftliche Datenerhebung oft eine Selbstentlarung des Untersuchungsobjektes voraus, im Rahmen derer „disfunktionale‘, normabweichende oder andere sanktionierte Eigenschaften öffentlich“ (Esser (1975b), S. 266) werden. Vermieden werden sollen demnach Nachteile, die sich aus der Angabe eindeutig als abweichend geltender Eigenschaften oder einer unerwarteten bzw. nicht abgesprochenen Veränderung des inhaltlichen Bezugsrahmens der Interaktionssituation ergeben. Hauptstrategie des Untersuchungsobjektes ist eine bewusste Sanktionsvermeidung durch Falschangaben. Hierbei ist allerdings zu beachten, dass das Antwortverhalten der Untersuchungsobjekte in einem solchen Fall als graduelle Abweichung vom wahren Wert zu verstehen ist und dabei von den Zielen und Interessen der Untersuchungsobjekte vor dem Hintergrund der jeweiligen Situationsdefinition abhängig sind. So ist die Auswahl einer adäquaten Reaktion, die den subjektiven Zielen des Untersuchungsobjektes dient, an eine Analyse der existenten Situationselemente und der aus diesen potentiell resultierenden Konsequenzen gebunden.

ESSER modelliert systematische intentionale Reaktivität in der Folge als Funktion situationaler Bedingungen. Eine systematische Reaktion ergibt sich demnach dann, „wenn hohe Nutzenintensitäten auf deutlich strukturierte subjektive Erwartungen über spezifische Folgen bestimmter Handlungen treffen“ (Esser (1986), S. 326). Hierzu muss zunächst die Möglichkeit zur Identifikation der Situationselemente und ihres Zusammenspiels gegeben sein („Sichtbarkeit“). Danach müssen sich aus „den (identifizierten) Situationsmerkmalen klar typisierte, unterschiedliche Konsequenzerwartungen“ (Esser (1986), S. 326) ergeben („Stereotypisierung“). Zuletzt muss die Situation eine gewisse Relevanz aufweisen, welche aus dem Grad der jeweiligen „Öffentlichkeit“ resultiert. Eine Verzerrung im Antwortverhalten ergibt sich dann als systematische Differenz der situationsspezifischen Reaktion zum wahren Wert. Voraussetzung hierfür ist allerdings, dass auch ein Unterschied zwischen beidem besteht. (Esser (1986), S. 326ff)

Reaktivität stellt ein im Grunde alle, aber im Besonderen sozialwissenschaftliche und naturwissenschaftliche¹, Disziplinen übergreifendes und auch die reale Lebenswelt betreffendes Problem dar. Dabei ist insbesondere die Validität der Forschungsergebnisse von Fachrichtungen, die auf empirische Daten zurückgreifen, durch reaktive Effekte bedroht. Zwar gelten innerhalb der empirischen Sozialforschung mit den verschiedenen Formen des Interviews nach wie vor Selbstbeurteilungs- bzw. Selbstauskunftsmaße als Königsweg der Datenerhebung, allerdings existiert mittlerweile auch eine Reihe von Ansätzen zur Lösung des Reaktivitätsproblems, bei denen es sich im Wesentlichen um sogenannte indirekte Messverfahren (physiologische Messungen, nicht-reaktive Messverfahren, implizite Maße, projektive Verfahren und Objektive Tests) handelt. Nachfolgend sollen nun die sogenannten direkten, reaktiven bzw. expliziten Messverfahren den sogenannten indirekten Messverfahren gegenübergestellt und deren jeweiligen Vor- und Nachteile diskutiert werden.

2.2 Direkte und indirekte Messverfahren

Neben der Unterscheidung des qualitativen und quantitativen Paradigmas existiert im Rahmen des Instrumentariums der empirischen Sozialforschung mit der Differenzierung von direkten und indirekten Messverfahren eine weitere grundlegende Distinktionsebene. Die beiden genannten Trennlinien verlaufen dabei quer zueinander, sodass sowohl in der qualitativen als auch in der quantitativen Forschungstradition direkte und indirekte Messverfahren zum Einsatz kommen. Es handelt sich hierbei allerdings um keine besonders trennscharfe Kategorisierung. Vielmehr sind die Übergänge zwischen den beiden Verfahrensklassen vergleichsweise fließend. (Fritzsche/Linneweber (2006), S. 191ff) Besonders deutlich wird dies anhand der Klassifizierung von Messmethoden als reaktives oder nicht-reaktives bzw. explizites oder implizites Verfahren. (Diekmann (2011), S. 629)

„Zunächst einmal ist die Dichotomisierung in reaktive und nichtreaktive Verfahren von der Dimension quantitativ versus qualitativ unabhängig. D.h., immer dann, wenn die qualitativen Verfahren so eingesetzt werden, daß keine Reaktivität gemäß der Definition von Campbell entsteht, wird dieses Verfahren zugleich ein nichtreaktives sein. Im Sinne eines solchen Vier-Felder-Schemas geht also, rein formal gesehen, die Trennungslinie quer durch eine Vielzahl der eingesetzten Verfahren.“ (Bungard/Lück (1995), S. 201)

So werden implizite Messverfahren an einigen Stellen als besondere Klasse nicht-reaktiver Verfahren verortet. Dies geschieht dabei mit Verweis darauf, dass sie für reaktive Effekte weitestgehend unempfindlich sein sollen. (Bortz/Döring (2016), S. 351; Jonas/Stroebe/Hewstone (2007), S. 63f) Hierbei handelt es sich aber allenfalls um ein notwendiges allerdings keinesfalls hinreichendes Kriterium. Tatsächlich erfordern alle impliziten Messverfahren eine Reaktion der Probanden. Anders als bei nicht-reaktiven Verfahren sind die Probanden somit aktiv in den Messprozess involviert. Die Gemeinsamkeit mit nicht-reaktiven Verfahren besteht allerdings darin, dass sich die Probanden des Inhalts der vorgenommenen Messung nicht bzw. weitestgehend nicht bewusst sind. Implizite Messverfahren enthalten somit Elemente reaktiver als auch nicht-reaktiver Messverfahren und bilden deshalb eine besondere Klasse indirekter Maße. Eine ähnliche Argumentation lässt sich für projektive Verfahren und Objektive Tests formulieren.

¹ Vgl. Heisenberg'sche Unschärferelation (Mühlenfeld (2004), S. 22)

„The qualification ‚direct/indirect‘ refers to the measurement procedure whereas the qualification ‚implicit/explicit‘ refers to the functional properties of the outcome of the measurement procedure. Each indirect measure produces an outcome for which it needs to be determined whether it can be characterized as implicit.“ (De Houwer (2006), S. 21)

Die Klassifikation als direktes bzw. indirektes Messverfahren ist vornehmlich an das Bewusstsein der Probanden für den Messprozess und im Besonderen dem Bewusstsein für dessen inhaltliche Komponente gebunden.

Direkte Messverfahren sind mit den sogenannten expliziten oder reaktiven Maßen gleichzusetzen und setzen somit notwendigerweise die Bereitschaft und Fähigkeit der Probanden zur Selbstauskunft bzw. -beurteilung voraus, was reaktiven Effekten Vorschub leisten kann.

„Whether a measure is direct or indirect is an objective property of the measure: If the measure involves asking participants to express the attitudes or thoughts that are measured, it is a direct measure. If the attitude or cognition is inferred from behavior other than a self-assessment of the participant, it is an indirect measure. There is no need to do research about this. The direct or indirect nature of the measure is determined simply by looking at its objective properties.“ (De Houwer (2006), S. 20)

Im Rahmen direkter Messverfahren sind sich die Probanden somit bewusst, Gegenstand eines Messprozesses zu sein, bzw. solche Messverfahren setzen eine bewusste Aufmerksamkeit voraus. Den Probanden ist das abgefragte Konstrukt bekannt und sie werden zu einer unmittelbaren Reaktion angehalten. Mit anderen Worten: Indirekte Messverfahren erfordern eine explizite Artikulation der interessierenden Informationen seitens der Probanden.

„With *direct* procedures, a person is asked to provide a self-report of his or her attitude. With *indirect* procedures an attempt is made to measure a person’s attitude without the person knowing it. In some cases, the person may think that something other than an attitude is being measured, and in other cases the measurement may be completely unobtrusive.“ (Petty/Cacioppo (1996), S. 9)

Indirekte Messverfahren schließen physiologische, nicht-reaktive, implizite und projektive Verfahren sowie sogenannte Objektive Tests ein. Kennzeichen dieser Verfahren ist der Versuch reaktive Effekte zu verhindern, indem je nach Verfahrenskategorie verschiedene der nachfolgenden Maßnahmen umgesetzt werden: Die Vermeidung von Selbsteinschätzungen bzw. die Verschleierung der Messintention durch indirekte Abfragen, die Ableitung von Erkenntnissen bezüglich der Zieldimension aus dem Verhalten der Probanden bzw. anhand der Erfassung verwandter Konstrukte und/oder Korrelate, die Minimierung der Interaktion zwischen Forscher und Proband bzw. die Billigung allenfalls indirekten Kontaktes zwischen beiden, die Datenerhebung innerhalb der natürlichen Lebenswelt der Probanden bzw. außerhalb von künstlich herbeigeführten Situationen und die Untersuchung eines Zielkonstrukts, ohne dass diese den Probanden bewusst ist bzw. die Probanden eine willentliche und absichtliche Anpassung ihres Antwortverhaltens vornehmen können.

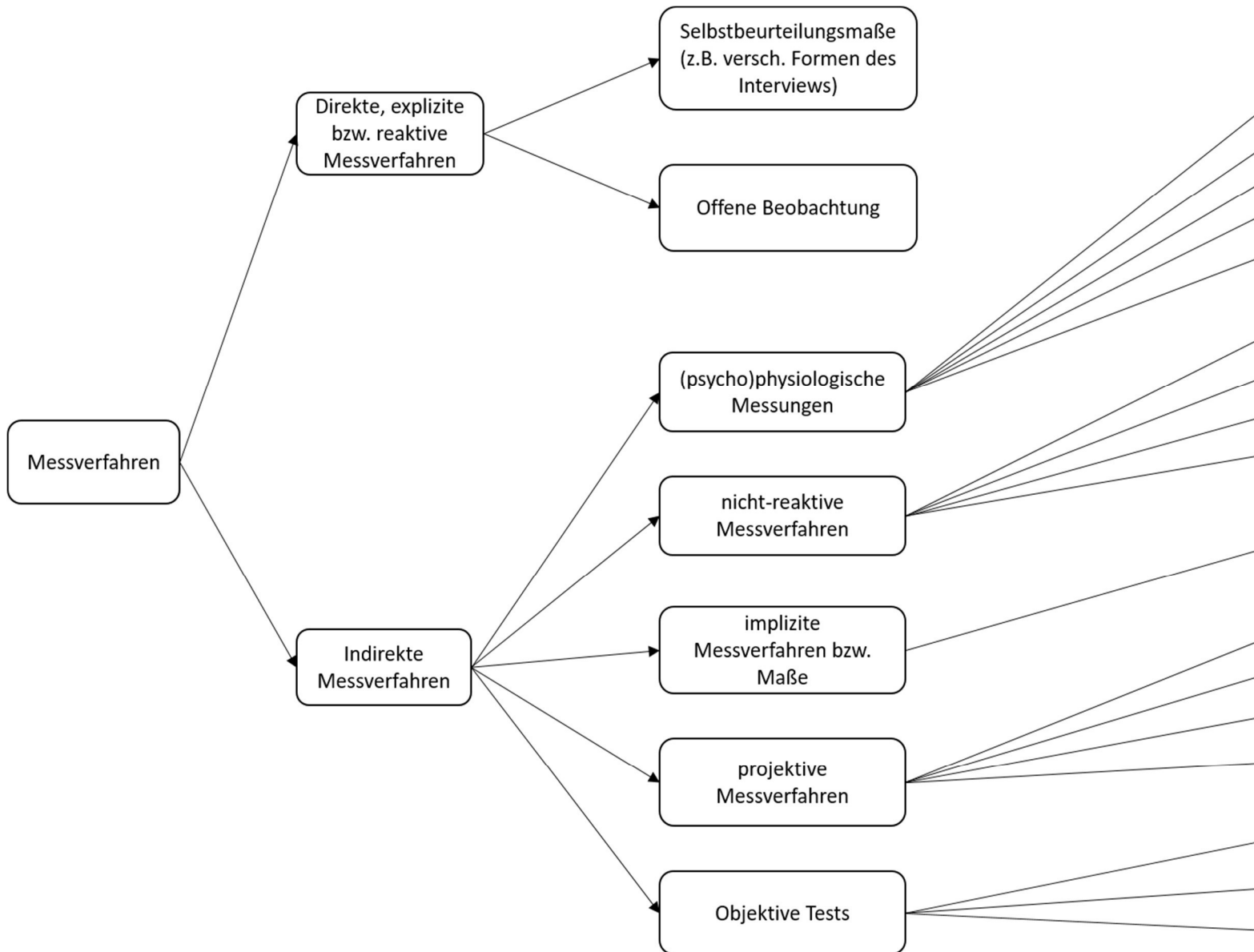


Abb 2.1: Übersicht über direkte und indirekte Messverfahren (Quelle: eigene Darstellung)

2.2.1 Direkte, explizite bzw. reaktive Messverfahren

Direkte Messverfahren sind im Wesentlichen mit sogenannten Selbstbeurteilungs- bzw. Selbstauskunftsmaßen, d.h. den verschiedenen Formen des Interviews, gleichzusetzen. (De Houwer (2003), S. 219; Gawronski/De Houwer (2014), S. 283; Petermann/Noack (1999), S. 440, S. 442; Petty/Cacioppo (1996), S. 9ff; Petty/Fazio/Briñol (2009), S. 3f) Eine Minimaldefinition kann daher lauten:

„Direct attitude measures are those that simply ask respondents to report their attitudes. Because these measures are transparent and make it obvious that attitudes are being assessed, they can be considered *explicit* measures. Included in this category are attitude measurement devices such as the semantic differential, the Likert scale, the Thurstone scale, and the ubiquitous one-item rating scale.“ (Petty/Fazio/Briñol (2009), S. 4)

„Specifically, a measurement procedure can be described as direct when the measurement outcome is based on participants' self-assessment of the to-be-measured attribute.“ (Gawronski/De Houwer (2014), S. 284)

Direkte Messmethoden enthüllen demnach ausdrücklich ihre Messintention und erfordern eine Selbstoffenbarung seitens der Probanden. Damit entsprechen ihre Merkmale denen von sogenannten expliziten Messverfahren. Auch explizite Messverfahren lassen sich im Wesentlichen als Selbsteinschätzungsmaße charakterisieren:

„(...); explicit methods (e.g. trait measures) are based on self-report and therefore require self-insight.“ (Robinson/Neighbors (2006), S. 115)

Zusätzlich setzen sie eine bewusste Aufmerksamkeit seitens der Probanden für die interessierende Zieldimension voraus und erwarten eine unmittelbare Artikulation von Informationen. Darüber hinaus bieten sie die Möglichkeit einer spezifischen Nachfrage und weisen eine für die Probanden nachvollziehbare und ersichtliche Augenscheinvalidität auf.

„Auf einem elementaren Niveau sind explizite Einstellungsmaße solche, bei denen die Befragten direkt gebeten werden, über ihre Einstellungen nachzudenken und darüber zu berichten (...).“ (Jonas/Stroebe/Hewstone (2007), S. 207)

„Conversely, measurement outcomes may be described as explicit if the impact of the to-be-measured psychological attribute on participants' responses is intentional, resource-dependent, conscious or controllable.“ (Gawronski/De Houwer (2014), S. 284)

Antworten auf explizite Messverfahren entsprechen demnach einer steuerbaren, intentionalen und bewussten kognitiven Reaktion seitens der Probanden. (Nosek (2007), S. 65) Direkte bzw. explizite Messverfahren verfügen somit auch über die Kennzeichen sogenannter reaktiver Messverfahren. Letztere zeichnen sich dadurch aus, dass sich die jeweiligen „Untersuchungsobjekte weitestgehend bewusst darüber sind, dass sie Gegenstand einer Untersuchung sind und die Möglichkeit haben, auf den Datenerhebungsvorgang selbst zu reagieren“ (Schnell/Hill/Esler (2011), S. 353).

„Reaktiv heißt, dass nichtkontrollierte Merkmale des Messinstruments, des Anwenders des Messinstruments (Verhalten des Versuchsleiters, des Interviewers) oder der Untersuchungssituation das Ergebnis der Messung

2. Theorie: Methodik

systematisch beeinflussen können. Kurz und einfach formuliert: Erhebungsmethoden sind reaktiv, wenn die Gefahr besteht, dass der Messvorgang das Messergebnis beeinflussen und verfälschen kann.“ (Diekmann (2011), S. 627)

„A reactive measure is one which modifies the phenomenon under study, which changes the very thing that one is trying to measure. In general, any measurement procedure which makes the subject self-conscious or aware of the fact of the experiment can be suspected of being a reactive measurement. Whenever the measurement process is not a part of the normal environment it is probably reactive. Whenever measurement exercises the process under study, it is almost certainly reactive.“ (Campbell (1957), S. 298f)

Zusätzlich lassen sich auch reaktive Messverfahren im Wesentlichen als Selbsteinschätzungsmaße verorten. (Petermann/Noack (1999), S. 442)

Direkte, explizite und reaktive Messverfahren scheinen somit wechselseitig über die gleichen Attribute zu verfügen und beschreiben schlussendlich sogenannte Selbstbeurteilungs- bzw. Selbstauskunftsmaße. Es lässt sich daher folgern, dass direkte Messverfahren mit expliziten bzw. reaktiven Messverfahren identisch sind.

Direkte Messverfahren in Form von Selbstbeurteilungs- bzw. Selbstauskunftsmaßen stellen die primäre und meistgenutzte Quelle sozialwissenschaftlicher Daten dar, nichtsdestotrotz sind mit ihnen sowohl Vorteile als auch Nachteile verbunden. Die Vorteile direkter Verfahren liegen in ihrer einfachen, schnellen, kostengünstigen und flexiblen Anwendbarkeit. Dabei sind sie in der Lage eine Vielzahl sozialer Aspekte, wie Einstellungen, Gefühle oder Verhalten, zu erfassen und weisen ungeachtet der mit Selbsteinschätzungen verbundenen Fehlerpotentiale eine vergleichsweise hohe Genauigkeit, Vorhersagekraft und Nützlichkeit auf. Die Validität und Reliabilität direkter Messverfahren ist nicht zuletzt deshalb als relativ hoch zu bewerten. Die Möglichkeit zur spezifischen Nachfrage ermöglicht darüber hinaus eine präzise Untersuchung sozialer Tatbestände. (Lucas/Baird (2006), S. 29ff, S. 41f; Petty/Cacioppo (1996), S. 22; Schwarz (1999), S. 93ff)

Direkte Messverfahren kommen dann an ihre Grenze, wenn Informationen bezüglich der Zieldimension nicht artikuliert werden können, weil beispielsweise kein Bewusstsein dafür vorhanden ist, einem Gegenstand gegenüber eine bestimmte Einstellung zu haben. (Jonas/Stroebe/Hewstone (2007), S. 209) Auch die Bereitschaft der Probanden zur aktiven Artikulation ihrer Einstellungen und Verhaltensweisen spielt bei der Anwendung direkter Messverfahren eine Rolle. Nur wenn diese Bereitschaft gegeben ist, können mit Hilfe dieser Methoden überhaupt Abfragen getätigt werden. Darüber hinaus können erhebliche Einschränkungen hinsichtlich der Reliabilität und Validität direkter Messverfahren aus deren methodischer Vorgehensweise erwachsen.

Zunächst müssen die Probanden die sogenannten *demand characteristics* bestimmen. Das heißt: Die Probanden müssen die ihnen gestellte Frage verstehen und interpretieren. Kommt es hierbei zu Fehleinschätzungen hat dies negative Konsequenzen für alle nachfolgenden Untersuchungsschritte. Wird beispielsweise eine von der Messintention des Forschers abweichende Zieldimension angenommen, führt dies zu erheblichen Verzerrungen im Messergebnis bzw. der Deutung des Messergebnisses. Je nachdem, ob alle Untersuchungsteilnehmer oder nur eine Stichprobe das interessierende Konstrukt fehlerhaft bestimmen, hat dies entweder einen massiven Bias oder mehr oder weniger große Abweichungen in den Resultaten zur Folge. (Bungard (1984), S. 116ff; Lucas/Baird (2006), S. 29, S. 31ff; Orne (1969), S. 147ff; Schwarz (1999), S. 94f)

Eine weitere Fehlerquelle ergibt sich aus dem Abruf der für die Beantwortung der Frage notwendigen bzw. relevanten Informationen aus dem Gedächtnis. Hier können

Erinnerungslücken, fehlerhafte Verknüpfungen von Ereignissen oder eine mangelnde Motivation, sich Verhalten, Einstellungen oder Ereignisse ins Gedächtnis zu rufen, zu Fehleinschätzungen führen. Im Falle von latenten Variablen, wie Einstellungen, Überzeugungen und Intentionen, müssen zudem erst Repräsentationen bzw. Indikatoren, in denen sich das jeweilige Konstrukt manifestiert, gefunden und evaluiert werden, bevor eine entsprechende Artikulation stattfinden kann. (Lucas/Baird (2006), S. 34ff)

Zuletzt weisen direkter Messverfahren eine erhebliche Sensitivität für sogenannte reaktive Effekte auf. (Jonas/Stroebe/Hewstone (2007), S. 207ff; Lemon (1973), S. 116ff; Schwarz (1999), S. 97) Reaktive Effekte stellen in diesem Sinne eine Reaktion des Befragten auf das Messinstrument, den Messkontext oder die Messsituation dar, in deren Folge der Befragte wissentlich oder unwissentlich eine Manipulation seines Antwortverhaltens vollzieht. Folge solcher Effekte ist in der Regel eine Diskrepanz zwischen wahrhafter Realität und erhobener Realität bzw. ein methodisches Artefakt. (Schnell/Hill/Esner (2008), S. 353; Scholl (2013), S. 85)

Reaktivitätseffekte ergeben sich aus der Kontext- und Situationsabhängigkeit des Messvorgangs und lassen sich nach SCHOLL in zwei Klassen unterscheiden. Im Fokus steht dabei die Befragung als Erhebungsmodus. Die genannten Effekte lassen sich allerdings auch auf andere Erhebungstechniken anwenden:

- (1) *Kognitive Effekte* hängen vom Verständnis von Frage- und Antwortvorgaben ab. So formt sich die Antwort eines Befragten erst im Zuge des Messvorgangs und orientiert sich dabei auch an Zusatz- und Kontextinformationen, die aus dem jeweiligen Messinstrument abgeleitet werden können und einen Hinweis auf die Art und Weise der intendierten Antwort liefern. Verzerrungen des wahren Wertes ergeben sich in diesem Fall aus der Frageformulierung, den zur Verfügung gestellten Antwortmöglichkeiten und der relativen Reihenfolge der Items. (Scholl (2013), S. 88; Scholl (2015), S. 211ff)
- (2) *Soziale Effekte* erwachsen aus der Interaktion von Versuchsleiter und Proband. Sie lassen sich im Wesentlichen in Versuchsleitereffekte, Antworttendenzen (*response sets*) bzw. Antwortstile (*response styles*) und Verfälschung differenzieren. So kann der Versuchsleiter bzw. Interviewer vermittelt durch persönliche Merkmale, wie bspw. Geschlecht, Alter, ethnische oder sozioökonomische Zugehörigkeit, den Interaktionsstil bzw. das Verhalten während des Messvorgangs sowie durch Erwartungen an das Messergebnis bzw. durch die Rolle des Probanden und Einstellungen gegenüber dem Forschungsobjekt Einfluss auf das Antwortverhalten der Probanden bzw. Befragten nehmen. Der Proband bzw. Befragte möchte im Rahmen der Messsituation wahrheitsgemäße Antworten geben, als kompetenter Gesprächspartner erscheinen und einen möglichst positiven Eindruck hinterlassen, sodass soziale und persönliche Anerkennung resultiert. Um diesen Zielen möglichst gerecht zu werden, werden die Konsequenzen verschiedener Antwortalternativen wechselseitig abgewogen, sodass ein optimales das Antwortverhalten nicht mehr unbedingt dem wahren Wert entspricht. (Möhring/Schlütz (2010), S.51ff, S. 56f; Scholl (2013), S. 89ff; Scholl (2015), S. 211f, S. 218ff)

2.2.1.1 Kognitive Reaktivitätseffekte

Messwertverzerrungen, die sich infolge der Informationsverarbeitung von Frage- und Antwortvorgaben ergeben, werden als kognitive Reaktivitätseffekte bezeichnet. Nachfolgend sollen einige Vertreter dieser Klasse von reaktiven Effekten vorgestellt werden.

- (1) Unterschiedliche sprachliche Formulierungen bzw. linguistische Differenz von Fragebogenitems können mitunter verschiedene Messergebnisse erzeugen. So bewirken oft eine bestimmte Wortwahl oder einzelne Schlüsselreizworte eine Konditionierung des Frageverständnisses oder üben Einfluss auf die Aktivierung von Gedächtnisinhalten und deren Verknüpfung mit Kontextinformationen aus. Hierbei ist allerdings zu beachten, dass Fragen grundsätzlich durch die Befragten gedeutet und in einen inhaltlichen Zusammenhang gebracht werden, der auch von der tatsächlichen Messintention abweichen kann. (Bungard (1984), S. 116ff; Clark/Schober (1992), S. 27ff; Lamp (2001); Petersen (2002), S. 234ff, S. 239f; Scholl (2015), S. 211f, S. 218ff)
- (2) Antwortvorgaben werden von den Probanden ebenfalls als Kontextinformationen wahrgenommen, wobei die jeweilige Skalenmitte dann zur Einordnung eines durchschnittlichen Verhaltens dient. So bewirken Ratingskalen, die im oberen bzw. unteren Skalenbereich differenzieren, jeweils eine Verzerrung des wahren Wertes nach oben bzw. nach unten. (Scholl (2015), S. 214f) Auch kann die Vorgabe spezifischer Antwortalternativen bzw. der Ausschluss einzelner Antwortalternativen eine bestimmte Manifestation des Forschungsgegenstandes bewirken. Die bereitgestellten Kategorien formen oder verändern dabei das Antwortverhalten der Befragten, sodass die Zieldimension nicht bloß beschrieben, sondern erst konstruiert wird. (Lemon (1973), S. 119)

Messwertverzerrungen, die aus der relativen Position der Items zueinander resultieren, werden als *Reihenfolgeeffekte* bezeichnet. Diese Ausstrahlung einer Frage auf weitere Fragen wird auch *Halo-Effekt*² genannt und hat eine Vielzahl von Ausprägungen. (Bradburn (1988), S. 302ff; Bühner (2011), S. 129f; Diekmann (2011), S. 464f; Scholl (2015), S. 216ff; Schumann/Presser (1981), S. 23ff; Strack (1992), S. 23ff; Sudman/Bradburn/Schwarz (1996), S. 263f) Beantworten die Probanden Items übereinstimmend, von denen sie glauben, dass sie das gleiche Merkmal erfassen bzw. wird durch eine erste Frage die Antwortrichtung für die nachfolgenden Fragen festgelegt, so spricht man vom *Konsistenz- bzw. Assimilationseffekt*. Dieser Effekt fällt unterschiedlich stark aus je nach Spezifität der Anfangsfrage, wobei die Festlegung auf eine Antwort eine logische Rationalisierung in Gang setzt. So wird die Antwort auf eine allgemeine Frage als Durchschnitt der Antworten auf die zugehörigen spezifischen Fragen angesehen. (Moosbrugger/Kelava (2012), S. 68; Petersen (2002), S. 226; Scholl (2015), S. 216) Der *Kontrasteffekt* lässt ein Item als Gegenpol zum nachfolgenden Item erscheinen. In der Folge ergibt sich eine stärkere Differenz im Antwortverhalten auf den beiden Items, wie wenn die erste Frage nicht gestellt worden wäre. Verzerrungen dieser Art ergeben sich aus einem Differenzierungsbedürfnis zwischen allgemeinen und spezifischen Fragen, einer Vermeidungstendenz gegenüber Antwortwiederholungen oder der Aktivierung verschiedener Grundstimmungen durch die jeweiligen Frageinhalte. (Brosius/Haas/Koschel (2012), S. 87; Petersen (2002), S. 226f; Scholl (2015), S. 216f) Kann ein nachfolgendes Item nicht unabhängig von einem vorangegangenen Item bearbeitet werden bzw. sind zur Beantwortung des nachfolgenden Items Informationen aus dem vorangegangenen Item notwendig, stellt dieses einen *Aktualisierungseffekt* dar. (Moosbrugger/Kelava (2012), S. 68) Werden Fragen am Ende eines Fragebogens weniger reflektiert beantwortet, spricht man vom „*fatigue effect*“. Sein

² Der *Halo-Effekt* kann auch als Urteilsfehler interpretiert werden und ist als solcher eher den sozialen Reaktivitätseffekten zuzuordnen. Er beschreibt dann „eine Tendenz, die Beurteilung mehrerer Merkmale eines Objektes von einem globalen Pauschalurteil abhängig zu machen, die Unfähigkeit oder mangelnde Bereitschaft des Urteilers, auf unterschiedliche Ausprägungen verschiedener Merkmale zu achten, oder die Tendenz eines Urteilers, ein Objekt bzgl. vieler Merkmale gleich einzustufen“ (Bortz/Döring (2016), S. 253).

Pendant, d.h. Fragen am Anfang eines Fragebogens werden weniger reflektiert beantwortet, wird als „*rapport effect*“ bezeichnet. Ein größeres Einverständnis mit Items am Anfang eines Fragebogens bzw. die Präferenz für zuerst genannte Antwortmöglichkeiten stellt einen „*primacy effect*“ dar, während ein größeres Einverständnis mit Items am Ende eines Fragebogens bzw. die Präferenz für letztgenannte Antwortmöglichkeiten einen „*recency effect*“ beschreiben. Effekte dieser Art ergeben sich daraus, dass die Aufmerksamkeit seitens der Probanden zu Beginn eines Fragebogens bzw. für am Anfang stehende Antwortmöglichkeiten hoch ist, danach abnimmt und zum Ende des Fragebogens bzw. der Auflistung wieder ansteigt. Items in der Mitte eines Fragebogens bzw. die mittleren Antwortvorgaben werden deshalb weniger reflektiert beantwortet bzw. weniger häufig ausgewählt. (Scholl (2015), S. 217f; Sudman/Bradburn/Schwarz (1996), S. 130)

2.2.1.2 Soziale Reaktivitätseffekte

Verzerrende Einflüsse, die aus der sozialen Situation des Messvorgangs bzw. aus der wechselseitigen Beziehung von Versuchsleiter und Proband respektive des Kommunikationsprozesses zwischen beiden resultieren, werden den sozialen Reaktivitätseffekten zugeordnet. Sie beschreiben offensive und defensive Strategien der Probanden, deren Ziel es ist, entweder eine von der Wirklichkeit abweichende Außendarstellung zu konstruieren, wobei eine situations- bzw. gegenstandsspezifisch möglichst positive Selbstpräsentation etabliert werden soll, oder die Determinierung einer konkreten Außendarstellung zu vermeiden. Auch innerhalb dieser Klasse reaktiver Effekte existieren verschiedene Abwandlungen, die nachfolgend exemplarisch vorgestellt werden sollen. Als Prototyp dieser Klasse reaktiver Effekte kann der sogenannte *Hawthorne-Effekt* herangezogen werden, wonach das Bewusstsein Gegenstand einer wissenschaftlichen Untersuchung zu sein bzw. sich in einer experimentellen Situation zu befinden, die Versuchspersonen dazu veranlasst, ihr Verhalten zu ändern. (Bortz/Döring (2016), S. 101; Diekmann (2011), S. 341)

- (1) *Versuchsleitereffekte* ergeben sich aus den sichtbaren und unsichtbaren Merkmalen der Interviewer. So können Geschlecht, Alter, Hautfarbe, ethnische Zugehörigkeit, Religionszugehörigkeit, sozioökonomischer Status und Auftreten des Versuchsleiters einen Effekt auf das Antwortverhalten der Probanden haben. Vom *Versuchsleiter-Erwartungseffekt* oder *Rosenthal-Effekt* spricht man, wenn die Probanden die Ergebniserwartungen des Versuchsleiters antizipieren und ihr Antwortverhalten entsprechend anpassen. Auf diese Weise ergibt sich eine Bestätigung der der jeweiligen Studie zugrundeliegenden Hypothesen im Sinne einer *self-fulfilling prophecy*. Ebenfalls eine Rolle spielen sogenannte *Sponsorship-Effekte*, bei denen sich die Kenntnis des Auftraggebers der Studie auf das Antwortverhalten der Probanden auswirkt. (Bortz/Döring (2016), S. 101; Bröder (2011), S. 160ff; Möhring/Schlütz (2010), S. 49ff; Huber (2009), S. 184ff; Shadish/Cook/Campbell (2002), S. 78f; Scholl (1993), S. 43ff)
- (2) *Formale Antwortstile* stellen einen Oberbegriff für reaktive Effekte dar, die sich dadurch auszeichnen, dass unabhängig vom Inhalt der gestellten Fragen konsequent ein bestimmtes Antwortverhalten gezeigt wird. So beschreiben die *Ja-Sage-* bzw. *Zustimmungstendenz* (*Akquieszenz*) und die *Nein-Sage-Tendenz* eine Antwortstrategie, bei der die Probanden auch bei inhaltlichem Widerspruch Fragen durchgängig positiv bzw. negativ beantworten. Ein solches Antwortmuster tritt häufig bei dichotomen Items auf (Bortz/Döring (2016),

S. 255; Bühner (2011), S. 129; Möhring/Schlütz (2010), S.58; Mummendey/Grau (2014), S. 156ff; Schnell/Hill/Esser (2008), S. 354f; Scholl (2015), S. 225) Die *Tendenz zur Mitte* meint eine Neigung der Probanden, grundsätzlich moderate Bewertungen vorzunehmen und extreme Antworten zu vermeiden. Ziel des Probanden ist es hierbei, sich nicht auf ein eindeutiges Urteil festlegen zu müssen. (Bortz/Döring (2016), S. 253f; Korman (1971), S. 180f; Scholl (2015), S. 225) Demgegenüber beschreibt der „*extremity bias*“ ein Antwortverhalten, bei dem durchgängig die äußeren Ausprägungen einer Antwortskala ausgewählt werden. Die Probanden vermeiden auf diese Art und Weise eine Differenzierung ihres Urteils. (Scholl (2015), S. 225; Möhring/Schlütz (2010), S.58) Der *Milde-Härte-Fehler* ist dadurch definiert, dass Objekte oder Personen systematisch zu positiv oder zu negativ bewertet werden. (Bortz/Döring (2016), S. 253) *Meinungslosigkeit* drückt sich durch die Auswahl der „*Weiß-nicht*“-Kategorie aus und ist dann als Verzerrung zu werten, wenn es sich um eine invalide Antwort handelt. Als ungültig gelten dabei *non-attitudes*, die aus einem Unverständnis der Frage sowie expliziter oder impliziter Antwortverweigerung resultieren. Die Probanden vermeiden an dieser Stelle einen kognitiven Aufwand, eine klare Positionierung bzw. Äußerung einer Präferenz sowie etwaige Nachfragen. (Brosius/Haas/Koschel (2012), S. 88f; Möhring/Schlütz (2010), S.107; Reuband (1990); Scholl (1993), S. 62ff; Scholl (2015), S. 225ff) *Pseudo-Meinungen* beschreiben die Äußerung einer Meinung, obwohl der jeweilige Proband sich noch keine Meinung gebildet hat. Hintergrund dieses Verhaltens ist es, der sozialen Norm entsprechend eine Meinung zu jedem Thema zu haben und Nachfragen zu vermeiden, die enthüllen könnten, dass dem nicht so ist. (Scholl (2015), S. 225)

- (3) *Verfälschbarkeit* bzw. *Verfälschung von Testwerten* beschreibt reaktive Effekte, die sich aus bewusst irreleitenden oder absichtlichen Falschaussagen seitens der Probanden ergeben. Ein solches Antwortverhalten zeigt sich vornehmlich bei der Abfrage besonders sensibler oder heikler Themen bzw. in Situationen, aus denen negative Konsequenzen für den Probanden folgen können. Daneben kann auch die Anwesenheit Dritter im Rahmen des Interviews oder das Interviewerverhalten Verfälschungstendenzen befördern. Auch sind insbesondere Items, die eine Selbstauskunft oder Selbstbeurteilung erfordern, durch Verfälschungstendenzen bedroht. Es lassen sich dabei zwei grundlegende Verfälschungsstrategien unterscheiden: (1) Simulation („*fake good*“) bezeichnet das Vortäuschen von Verhalten, das man sonst nicht zeigen würde, während (2) Dissimulation („*fake bad*“) das Verschleiern oder Verbergen von Verhalten beschreibt, das man normalerweise zeigen würde. (Bortz/Döring (2016), S. 436f; Bühner (2011), S. 125ff; Scholl (2013), S. 91)

2.2.1.2.1 Soziale Erwünschtheit

Sozialerwünschtes Antwortverhalten stellt eine spezielle Form der Verfälschungstendenz dar, die im Wesentlichen durch ein „*faking good*“, d.h. eine Beschönigungstendenz, gekennzeichnet ist. (Hartmann (1991), S. 64f) Soziale Erwünschtheit („*Social Desirability*“) beschreibt die Neigung der Probanden, sich in ihrem Antwortverhalten weniger an ihrem tatsächlichen Verhalten als an sozialen Normen zu orientieren, und tritt dabei vornehmlich in Situationen auf, in denen man sich besonders gut darstellen möchte, in denen negative Konsequenzen aus einem nicht-normkonformen Verhalten drohen oder in denen man ein positives Bild von sich erhalten

möchte. Dabei sollen positives Verhalten bzw. sozial konforme Verhaltensweisen oder besonders günstige Eigenschaften oder Merkmale hervorgehoben werden, während gleichzeitig als negativ empfundenes Verhalten bzw. normverletzende Verhaltensweisen oder unerwünschte Eigenschaften und Merkmale verborgen werden sollen. Welches Verhalten dabei als positiv oder negativ gilt, kann nicht universell bestimmt werden, sondern ist individuell, je nach Gruppe, situationsabhängig und kulturspezifisch verschieden. (Bortz/Döring (2016), S. 437ff; Bühner (2006), S. 60ff; Schnell/Hill/Esser (2008), S. 355f)

Soziale Erwünschtheit manifestiert sich entweder als Persönlichkeitsmerkmal in Form eines Strebens nach sozialer Anerkennung („*need for social approval*“) oder ist an die jeweils abgefragten Merkmale („*trait desirability*“) gebunden. In letzterem Fall ergibt sich sozialerwünschtes Antwortverhalten aus einem Bewusstsein darüber, dass innerhalb der Gesellschaft ein Konsens über wünschenswerte Einstellungen und Verhaltensweisen besteht. (Möhring/Schlütz (2010), S. 62; Mummendey/Grau (2014), S. 171f; Reinecke (1991), S. 94ff; Schnell/Hill/Esser (2008), S. 355; Scholl (2015), S. 220f)

Je nach Situation und Bedrohungsgrad der Frage ergeben sich Unterschiede hinsichtlich der Sensitivität für soziale Erwünschtheit. So spielen der Grad der Öffentlichkeit der Befragungssituation und die Tatsache, ob es sich um ein heikles oder sensibles Thema handelt, eine große Rolle. Ganz allgemein muss allerdings für aus sozialerwünschtem Antwortverhalten resultierende Verzerrungen eine Differenz zwischen wahrem Wert und erhobenem Wert existieren. (Möhring/Schlütz (2010), S. 62; Schnell/Hill/Esser (2008), S. 356; Scholl (2015), S. 221f)

Wie beschrieben gilt eine sozialerwünschte Antwort nur dann als unaufrichtig, wenn die Probanden das sozialerwünschte Verhalten nur vortäuschen bzw. ein tatsächlich praktiziertes sozial unerwünschtes Verhalten unterschlagen. SCHOLL differenziert in diesem Zusammenhang zwischen psychologischer und sozialer Unaufrichtigkeit. Psychologische Unaufrichtigkeit entspricht dabei einer bewussten Täuschung im Sinne einer punktuellen, situationspezifischen Lüge, während soziale Unaufrichtigkeit einer konsistenten, situationsübergreifenden Lüge entspricht. Hieraus ergeben sich vier idealtypische Konstellationen für das Antwortverhalten der Probanden (Scholl (2015), S. 219f):

- (1) Die Probanden antworten bezogen auf ein bestimmtes Thema sowohl im Rahmen der Erhebungssituation als auch in alltäglichen Situationen konsistent aufrichtig. Ein solches Antwortverhalten wird als valide betrachtet.
- (2) Die Probanden antworten ausschließlich im Rahmen der Erhebungssituation aufrichtig, während in Alltagssituationen die gleiche Thematik Falschauskünfte provoziert. Ein solches Antwortverhalten gilt als valide, solange das Ziel der Erhebung nur die Abfrage des tatsächlichen Verhaltens der Probanden beinhaltet.
- (3) Die Probanden antworten sowohl im Rahmen der Erhebungssituation als auch im Alltag gleichermaßen unaufrichtig. Dieses Antwortverhalten gilt als invalide, wenn ausschließlich die Erfassung einer gültigen Antwort im Vordergrund steht.
- (4) Die Probanden antworten ausschließlich in der Erhebungssituation unaufrichtig, während die gleiche Thematik im Alltag der Wahrheit entsprechende Antworten hervorruft. Ein solches Antwortverhalten ist als invalide zu bewerten.

Messtechnisch wird soziale Erwünschtheit als multidimensionales Konstrukt begriffen. Dabei hat sich ein zweifaktorielles Modell durchgesetzt, wonach das Phänomen soziale Erwünschtheit aus einer Selbsttäuschungs- („*self-deception*“) und einer Fremdtäuschungskomponente („*impression management*“) besteht. Dabei beschreibt Selbsttäuschung die Tendenz das

Selbstbild und das Selbstwertgefühl durch ein bestimmtes Antwortverhalten zu schützen und das eigene Selbst bzw. die Realität in einer optimistischen Art und Weise verzerrt wahrzunehmen. Fremdtäuschung meint dagegen eine bewusste Irreführung bzw. Verfälschung, um nach außen ein möglichst vorteilhaftes Bild abzugeben bzw. um eine absichtliche und zielgerichtete Täuschung anderer herbeizuführen. (Möhring/Schlütz (2010), S. 62; Musch/Brockhaus/Bröder (2002), S. 121; Paulhus (1984), S. 599, S. 605ff; Paulhus (1986), S. 144ff; Winkler/Kroh/Spiess (2006), S. 2f)

2.2.1.2.2 Kontrolle sozialerwünschten Antwortverhaltens

Es existiert eine Vielzahl von Verfahren bzw. Strategien, um den Einfluss sozialer Erwünschtheit auf das Antwortverhalten in Fragebögen zu kontrollieren. HERZBERG unterscheidet an dieser Stelle nach dem Zeitpunkt der Kontrollmaßnahme. (Herzberg (2011), S. 137ff)

(1) Testkonstruktion

- a.) *Itemkonstruktion und -selektion*: Durch die Art und Weise der Itemformulierung bzw. der Frageform kann bewirkt werden, dass die Probanden eher bereit sind, auch unerwünschtes Verhalten offenzulegen. Dies kann durch einen ausdrücklichen Hinweis auf die heikle Thematik der Frage geschehen. Abweichendes Verhalten wird mitunter schlichtweg unterstellt und nur nach der Häufigkeit des fraglichen Verhaltens gefragt. Abweichendes Verhalten wird als normal definiert. Unerwünschtes Verhalten wird als Ausnahme dargestellt. Es findet eine Entschärfung statt, indem Aussagen von Autoritäten in das Item miteingebunden werden. Es werden offene Fragen oder besonders lange Fragen verwendet, um eine individuelle Beschreibung des Verhaltens zu ermöglichen bzw. das Gedächtnis der Probanden zu stimulieren und auf diese Weise die Berichterstattung für die Probanden zu erleichtern. Das unerwünschte Verhalten wird in einen weniger bedrohlichen Kontext eingebettet bzw. an eine spezifische soziale Situation angeschlossen. Im Rahmen der *nominativen Technik* nominiert der Versuchsleiter eine Person aus dem Umfeld der Probanden, wobei angegeben werden soll, ob diejenige Person abweichendes Verhalten zeigt. (Möhring/Schlütz (2010), S. 64; Mummendey/Grau (2014), S. 175f; Scholl (2015), S. 223f)
- b.) *Forced-Choice-Items bzw. ausbalancierte Antwortvorgaben*: Den Probanden werden verschiedene Aussagen vorgelegt, wobei mindestens eine dieser Aussagen als Antwort ausgewählt werden muss. Voraussetzung ist, dass alle Antwortalternativen eines Items denselben Grad an sozialer Erwünschtheit aufweisen. Den Probanden bleibt somit keine Möglichkeit durch ihre Antwort einen bestimmten Eindruck vorzutäuschen. (Bortz/Döring (2016), S. 439; Bühner (2011), S. 127f; Mummendey (1981), S. 206; Mummendey/Grau (2014), S. 176f)
- c.) *Einsatz von Kontrollskalen*: Es werden Items bzw. sogenannte SD-Skalen in den Fragebogen aufgenommen, die besonders sensibel auf sozialerwünschtes Antwortverhalten reagieren. (Mummendey (1981), S. 206f; Mummendey/Grau (2014), S. 178ff)
- d.) *Objektive Tests*: Die Zieldimension bzw. das Testziel wird durch geschickte Aufgabenauswahl und Auswertungstechnik für die Probanden möglichst undurchschaubar gestaltet. (Bortz/Döring (2016), S. 438)

(2) Testadministration

- a.) *Random-Response-Technik*: Die geprüfte Person kann sich aufgrund eines speziellen Zufallsverfahrens absolut sicher sein, dass sich ihr Antwortverhalten nicht rekonstruieren lässt. Für jedes Item wird per Zufall entschieden (z.B. durch Würfeln), ob ehrlich geantwortet wird. Es wird nun angenommen, dass aufgrund der geringen Rückverfolgungschancen die Probanden nun keine Veranlassung mehr zur Verfälschung ihrer Antworten haben. (Bortz/Döring (2016), S. 438; Diekmann (2011), S. 488ff; Herzberg (2011), S. 140; Warner (1965))
- b.) *Bogus-Pipeline-Methode*: Während der Bearbeitung des Fragebogens sind die Probanden an einen vermeintlichen Lügendetektor angeschlossen, was in einer signifikanten Reduktion von Falschantworten resultiert. Voraussetzung für eine Reduktion sozialerwünschten Antwortverhaltens ist allerdings, dass den Versuchspersonen erfolgreich die potentielle Prüfung des Wahrheitsgehaltes ihrer Aussagen suggeriert werden kann. (Herzberg (2011), S. 140; Jones/Sigall (1971); Mummendey (1981), S. 208ff, Mummendey/Grau (2014), S. 187ff; Scholl (2015), S. 221)
- c.) *Fragebogen-Instruktion*: Der Befragung werden verschiedene Testinstruktionen vorangestellt, deren Befolgung eine Reduktion der Beschönigungstendenz zur Folge haben soll. Dies kann eine Aufforderung zu korrekter Testbearbeitung sein, welche die Probanden dazu anhält „offen und ehrlich“ zu antworten. Eine Tempoinstruktion weist die Probanden an, das Untersuchungsinstrument möglichst zügig zu bearbeiten, sodass kaum Zeit bleibt, die Antworten bewusst zu verfälschen. Eine Anti-SD-Instruktion weist die Probanden explizit daraufhin, dass Antworttendenzen unmittelbar erkannt werden und deshalb vermieden werden sollten. (Bühner (2011), S. 127; Mummendey (1981), S. 207; Mummendey/Grau (2014), S. 182ff)
- d.) *Faking-Good- vs. Faking-Bad-Instruktion*: Das Testinstrument wird von den Probanden unter der Bedingung, sich möglichst positiv darzustellen, sich möglichst negativ darzustellen, und unter normalen Bedingungen bearbeitet. Aus dem Vergleich des jeweiligen Antwortverhaltens lassen sich Rückschlüsse auf die Anfälligkeit des Untersuchungsinstrumentes für Verfälschungen ableiten. (Bortz/Döring (2016), S. 437; Mummendey (1981), S. 207)
- e.) *Anonymitätsgarantie*: Mündliche oder schriftliche Zusicherung, dass die Identität der Probanden nicht mit ihren Angaben verknüpft wird. (Mummendey/Grau (2014), S. 184)

(3) Testauswertung

- a.) *Faktor-Elimination*: Der im Rahmen einer Hauptkomponentenanalyse extrahierte erste Faktor enthält solche Items, die besonders viel gemeinsame Varianz aufweisen, und erklärt verglichen mit den weiteren Faktoren den größten Anteil an Varianz. Neben einem Varianzanteil, der auf inhaltliche Komponenten zurückzuführen ist, besitzen die Items, die auf den ersten Faktor laden, einen gemeinsamen Varianzanteil, der ausschließlich auf sozialerwünschtes Antwortverhalten zurückzuführen ist. Der größte Teil der erklärten Varianz ist dabei sozialer Erwünschtheit zuzurechnen, sodass dieser Faktor auch als sozialer Erwünschtheitsfaktor interpretiert werden kann. (Herzberg (2011), S. 138; Mummendey/Grau (2014), S. 181f)
- b.) *Kontrolle von Extremfällen*: Es werden Personen von der Analyse ausgeschlossen, die SD-Werte oberhalb eines kritischen Wertes aufweisen oder es erfolgt eine statistische Korrektur auf Basis der individuellen SD-Werte. (Herzberg (2011), S. 140)

2.2.1.3 Repräsentativität

Während kognitive und soziale Reaktivitätseffekte die Validität einer Messung beeinflussen und beeinträchtigen, existiert mit *Non-Coverage*- und *Unit-Nonresponse*-Fehlern eine weitere Klasse reaktiver Effekte, die vornehmlich die Repräsentativität einer Studie bedrohen. (Möhring/Schlütz (2010), S.42) Beide Fehlerarten betreffen die Teilnahme potentiell relevanter Fälle an empirischen Studien bzw. die Problematik, dass oft nicht alle Fälle, die für eine Stichprobe vorgesehen waren, auch tatsächlich in die Stichprobe aufgenommen werden konnten. *Non-Coverage* beschreibt dabei die prinzipielle Nicht-Erreichbarkeit bestimmter Personengruppen, während *Unit-Nonresponse* im Wesentlichen die bewusste Teilnahmeverweigerung beinhaltet. Ergebnisverzerrungen ergeben sich hierbei aus der Unvollständigkeit der Datengrundlage in Form ungenügender Varianz bzw. dadurch, dass möglicherweise relevante Fälle unberücksichtigt bleiben. (Schnell/Hill/Esser (2008), S. 306ff; Scholl (2015), S. 228ff)

2.2.2 Indirekte Messverfahren

Indirekte Messverfahren beschreiben Methoden, „die von der Selbsterkenntnis und von dem Willen der Person, sich zu enthüllen, in weitem Maß unabhängig sind“ (Sellitz/Jahoda/Deutsch/Cook (1972), S. 61). Sie können dabei als Alternative zu direkten Messverfahren verstanden werden und versuchen die mit dieser Methodenfamilie einhergehenden Probleme hinsichtlich reaktiver Effekte zu lösen.

„Indirect attitude measures, on the other hand, are those that do not directly ask the individual to report his or her attitude. Such measures were to be used when it was either impractical or undesirable to ask people what their opinions were (...), or there was some possibility that people might not be willing to tell you what their attitudes were (...). When using indirect measurement approaches, the individual's attitude is inferred from his or her judgements, bodily responses, or overt behaviors.“ (Petty/Fazio/Briñol (2009), S. 4)

Indirekte Messverfahren zeichnen sich dadurch aus, dass sie eine mittelbare Abfrage des interessierenden Konstruktes vollziehen, indem sie Selbsteinschätzungen seitens der Probanden versuchen zu vermeiden. Vielmehr wird anhand von Verhalten oder der Abfrage verwandter Konstrukte bzw. von Korrelaten Erkenntnis hinsichtlich des zu untersuchenden Gegenstandes abgeleitet. Indirekte Messverfahren können demnach auch Selbstbeurteilungs- bzw. Selbstauskunftsmaße enthalten, wobei der Aspekt, der abgefragt wird, nicht der eigentlichen Zieldimension entspricht, sondern einen indirekten Zugang zu dieser gewährt. Im Allgemeinen wird die Messintention für die Probanden möglichst undurchschaubar gestaltet, sodass die Probanden sich nicht bewusst sind, was gemessen werden soll, und sie keine willentliche Modifikation ihres Antwortverhaltens vornehmen können, die Verzerrungen im Messergebnis nach sich ziehen könnte. Indirekte Messverfahren entsprechen somit im Wesentlichen dem, was nachfolgend als Objektive Tests bezeichnet werden wird, umfassen allerdings noch weitere Subklassen, wie physiologische bzw. psychophysiologische Messungen, nicht-reaktive bzw. unaufdringliche Messverfahren, implizite Messverfahren bzw. Maße oder projektive Verfahren, welche nachfolgend vorgestellt werden sollen. (De Houwer (2006), S. 19ff; Petermann/Noack (1999), S. 440f; Petty/Cacioppo (1996), S. 9, S. 16ff; Schmukle/Egloff (2011), S. 73ff)

2.2.2.1 Physiologische bzw. psychophysiologische Messungen

Physiologische bzw. psychophysiologische Messungen werden in den Sozialwissenschaften hauptsächlich parallel zu anderen Methoden verwendet, wobei die jeweiligen Messergebnisse wechselseitig zueinander in Beziehung gesetzt werden, um Rückschlüsse auf psychische Prozesse ziehen zu können.

„Die physiologischen Messungen („physiological measurements“) dienen der objektiven Erfassung und Quantifizierung bestimmter Merkmale physiologischer Prozesse in unterschiedlichen Organsystemen des Körpers mittels entsprechender Messgeräte.“ (Bortz/Döring (2016), S. 501)

„Die psychophysiologischen Messungen („psychophysiological measurements“) entstehen, wenn man physiologische Messungen mit Merkmalen des Verhaltens und Erlebens zeitbasiert verknüpft.“ (Bortz/Döring (2016), S. 501)

Die Vorteile psychophysiologischer Maße liegen in ihrer, verglichen mit Selbstauskunftsmaßen, geringen Anfälligkeit für gezielte Verfälschungen, der Vermeidung von Verzerrungen infolge von Gedächtnisfehlern, der Möglichkeit sowohl nicht bewusst wahrgenommene als auch nicht eindeutig verbalisierbare Phänomene und Inhalte zu erfassen, die Dokumentation von Verlaufsdaten und dem Einblick in Prozesse der Informationsverarbeitung. (Bortz/Döring (2016), S. 502)

Ihre Nachteile betreffen den untersuchungstechnischen Aufwand hinsichtlich Material und Kosten, die Notwendigkeit von Expertenwissen für die Durchführung und Interpretation von Messungen, die aus Messaufbau und Messungsdauer resultierende potentielle Unzumutbarkeit für die Probanden sowie die Tatsache, dass nur indirekte Rückschlüsse auf subjektive Charakteristika der Probanden möglich sind. Da physiologische Messungen normalerweise unter Laborbedingungen stattfinden, sind sich die Probanden in der Regel bewusst, Teil einer wissenschaftlichen Untersuchung zu sein, und können gegebenenfalls auf diese reagieren. Daneben können auch physiologische und technische Artefakte auftreten, die durch bestimmte Verhaltensweisen oder Bewegungen der Probanden während des Messvorgangs ausgelöst werden oder durch technische Störeinflüsse, wie elektromagnetische Felder, Wackelkontakte oder defekte Messgeräte zustande kommen. (Bortz/Döring (2016), S. 502ff)

Außerdem sind methodische Probleme, wie die Spezifitäts- und die Ausgangswertproblematik zu beachten. Erstere beschreibt den Umstand, dass unabhängig vom Stimulus die Probanden grundsätzlich ein für sie typisches Reaktionsmuster zeigen. Dieser Mangel an Varianz in der Probandenreaktion lässt in der Folge keine differenzierten Zusammenhangsschlussfolgerungen zu. Letztere betrifft die Differenz von Verlaufs- und Ausgangswert (Baseline), die je nach gesetztem Stimulus und Höhe des Ausgangswertes unterschiedlich ausfallen kann. So erreichen beispielsweise Personen mit einem höheren Ausgangswert grundsätzlich niedrigere Differenzen, wenn der Stimulus eine Steigerung des Biosignals bewirken soll, und sie erreichen grundsätzlich höhere Differenzen, wenn der Stimulus eine Senkung des Biosignals bewirken soll. Für Personen mit einem niedrigen Ausgangswert verhält es sich umgekehrt. Bei der Bewertung einer solchen Differenz müssen somit auch unbedingt immer die individuellen Charakteristika der Teilnehmer beachtet werden. (Bortz/Döring (2016), S. 504; Schandry (1996), S. 550ff)

Typische physiologische Messverfahren betreffen die Aktivitäten des Gehirns, des Herz-Kreislauf-Systems, der Haut, der Muskeln und der Augen. Zur Erfassung und Analyse der

Hirnaktivität werden vornehmlich die Elektroenzephalografie (EEG) und die Magnetresonanztomografie (MRT) herangezogen. Im Fokus steht dabei die Untersuchung von Wahrnehmungs-, Verhaltens-, Denk- und Urteilsprozessen, Emotionen, Motivation sowie Handlungsplanung und -kontrolle. Im Rahmen der Untersuchung des Herz-Kreislauf-Systems werden anhand von Herzschlagfrequenz, Pulsfrequenz und Blutdruck bzw. den Ergebnissen der Elektrokardiografie (EKG) Rückschlüsse auf motorische Leistungsfähigkeit, Lern- und Problemlösungsprozesse, Aufmerksamkeitsprozesse oder Motivationen gezogen. Bei der elektrodermalen Aktivität steht die Messung des Hautwiderstandes bzw. der Hautleitfähigkeit im Fokus. Aus diesen lassen sich z.B. Informationen bezüglich der Reaktionsschnelligkeit, der emotionalen Erregung oder von Aufmerksamkeitsprozessen ableiten. Bei der Analyse der muskulären Aktivität stehen insbesondere die Mimik bzw. die Bewegungen der Gesichtsmuskulatur im Mittelpunkt. Anhand dieser lassen sich sowohl Emotionen, Einstellungen und affektive Störungen ablesen als auch Lügen detektieren. Augen vollziehen prinzipiell drei Arten von Bewegungen (Vergenz- und Folgebewegungen sowie sakkadische Bewegungen), die mittels Elektrookulografie und Eye-Tracking erfasst werden. Die Elektrookulografie gibt dabei anhand der elektrischen Aktivität der Augenmuskulatur Auskunft darüber, wie lange Personen wohin schauen. Eye-Tracking erfasst Blickbewegungen mit Hilfe hochauflösender Kameras. Die Untersuchung der Augenaktivität dient zur Erkenntnis bezüglich kognitiver Effekte, der Aufnahme visueller Informationen, psychischer Erkrankungen, Werbewirksamkeit und der Methodenforschung. (Bortz/Döring (2016), S. 499ff; Petermann/Noack (1999), S. 441; Petty/Cacioppo (1996), S. 18ff; Schandry (1996), S. 554ff)

2.2.2.2 Nicht-reaktive bzw. unaufdringliche Messverfahren

Die Entwicklung nicht-reaktiver bzw. unaufdringlicher Messverfahren geht auf WEBB/CAMPBELL/SCHWARTZ/SECHREST/GROVE zurück, die diese als Antwort auf die methodischen Schwächen von Interviews und Fragebogen betrachten. Die mit solchen Selbstbeurteilungs- bzw. Selbstauskunftsmaßen oder reaktiven Messverfahren einhergehenden methodischen Probleme fassen sie dabei wie folgt zusammen:

„Interviews and questionnaires intrude as a foreign element into the social setting they would describe, they create as well as measure attitudes, they elicit atypical roles and responses, they are limited to those who are accessible and will cooperate, and the responses obtained are produced in part by dimensions of individual differences irrelevant to the topic at hand.“ (Webb/Campbell/Schwartz/Sechrest/Grove (1981), S. 1)

Die Verwendung von Selbstbeurteilungs- bzw. Selbstauskunftsmaßen hat somit mitunter sogenannte reaktive Effekte zur Folge, die Verzerrungen innerhalb der erhobenen Daten nach sich ziehen, welche die Validität der Forschungsergebnisse bedrohen können. In letzter Instanz wird möglicherweise erst durch den Forschungsprozess die zu untersuchende Realität konstruiert oder der Forschungsgegenstand verändert. (Albrecht (1972), S. 243; Sechrest/Phillips (1979), S. 3f)

Nicht-reaktive Messmethoden versuchen Reaktivität und deren Konsequenzen für das Messergebnis dadurch auszuschalten, dass sich die Probanden der Teilnahme an einem Messprozess nicht bewusst sind. Um dies zu gewährleisten, werden im Wesentlichen zwei Strategien verfolgt:

2. Theorie: Methodik

(1) Zunächst kann angenommen werden, dass „the reactivity of a measure increases with the extent to which both the research subject and the researcher are involved in the act of measurement“ (Fritsche/Linneweber (2006), S. 191). Ziel nicht-reaktiver Verfahren ist es deshalb, die direkte und bewusste Interaktion zwischen Forscher und Proband zu eliminieren. Der Forscher tritt allenfalls indirekt mit dem Forschungsgegenstand in Kontakt. Auf diese Weise soll verhindert werden, dass der Forscher das Verhalten der Probanden beeinflussen kann. Nicht-reaktive Verfahren bedienen sich vielmehr aller Arten von Spuren, die durch das Verhalten der Probanden hinterlassen werden. (Berg (2009), S. 268; Bungard/Lück (1995), S. 202f; Diekmann (2011), S. 628f; Fritsche/Linneweber (2006), S. 190f; Mayrhofer (1993), S. 15) Eine vorläufige Definition nicht-reaktiver Messverfahren kann daher lauten:

„Nonreaktive Verfahren der Datenerhebung („nonreactive/nonintruding/unobtrusive measures“) umfassen allgemein alle Datenerhebungstechniken, bei denen das Verhalten und Erleben der Untersuchungsteilnehmenden nicht durch Interaktion mit den Forschenden beeinflusst werden kann. Die erhobenen Daten sind somit unverzerrt von möglichen Reaktionen der Untersuchten auf den Forschungsprozess selbst.“ (Bortz/Döring (2016), S. 349)

(2) Zusätzlich verzichten nicht-reaktive Messverfahren im Gegensatz zu reaktiven Messverfahren auf eine Datenerhebung in künstlich herbeigeführten Situationen. Vielmehr findet die Sammlung von Daten innerhalb der vertrauten bzw. alltäglichen Umgebung der Forschungsobjekte statt, sodass der Messvorgang Bestandteil der gewohnten Lebenswelt wird. Die Folge sind unverfälschte Reaktionen auf Alltagsbedingungen, ohne dass sich die Probanden bewusst sind, Daten für eine wissenschaftliche Untersuchung zu liefern. (Mayrhofer (1993), S. 15; Petermann/Noack (1999), S. 443) Entsprechend lautet eine abschließende Definition nicht-reaktiver Messverfahren wie folgt:

„Der Grundgedanke der nichtreaktiven Verfahren ist dabei der, daß man den Meßvorgang in die ‚normale‘ Umgebung der Probanden verlegt und auf eine direkte bzw. bewußte Interaktion zwischen Forscher, Interviewer bzw. Beobachter und Untersuchungsobjekt verzichtet. Die Datenerhebung erfolgt also unaufdringlich, unbemerkt, die Versuchsperson weiß nicht, daß sie an einer wissenschaftlichen Untersuchung teilnimmt.“ (Bungard/Lück (1995), S. 199)

WEBB/CAMPBELL/SCHWARTZ/SECHREST/GROVE fassen die Problematik reaktiver Effekte und die entsprechenden Gegenmaßnahmen nicht-reaktiver Verfahren wie folgt zusammen:

„This flexibility of thought is required to handle reactive measurement effects, which are the most systematic weakness of all interview and questionnaire studies. These error threats are also systematically present in all observation studies in which the presence of an observer is known to those under study. To varying degrees, measurements conducted in natural settings, without the individual's knowledge, control this type of error possibility. In all of them – hidden observation, contrived observation, trace analysis, and secondary records – the individual is not aware of being tested, and there is little danger that the act of measurement will itself serve as a force for change in behavior or elicit role playing that confounds the data. There is also minimal risk that biases coming from the physical appearance or other cues provided by the investigator will contaminate the results.“ (Webb/Campbell/Schwartz/Sechrest/Grove (1981), S. 316)

Nicht-reaktive Messverfahren sollten dabei allerdings weniger als Alternative zu reaktiven Messverfahren verwendet werden, sondern vornehmlich im Sinne einer Triangulation verschiedener Verfahren zu Erkenntnissen bezüglich des Forschungsgegenstandes beitragen, d.h. als Ergänzung zu reaktiven Messverfahren fungieren. (Albrecht (1972), S. 244; Berg (2009), S. 287; Lee (2000), S. 6f)

Nach WEBB/CAMPBELL/SCHWARTZ/SECHREST/GROVE sind die grundlegenden nicht-reaktiven Messverfahren die Analyse physikalischer Spuren bzw. von Verhaltensspuren, die inhaltsanalytische Untersuchung von Archivmaterial bzw. Dokumenten, die verdeckte Beobachtung und das Feldexperiment. (Sechrest/Phillips (1979), S. 9; Webb/Campbell/Schwartz/Sechrest/Grove (1981))

Die Analyse physikalischer Spuren befasst sich mit der Erfassung und Auswertung menschlicher Hinterlassenschaften, aus denen sich Informationen hinsichtlich der Verhaltensregelmäßigkeiten und Lebensgewohnheiten der Forschungsobjekte ableiten lassen. Solche Spuren wurden meist unbewusst und unbeabsichtigt von ihren Verursachern zurückgelassen, sodass sie in dem Sinne nicht-reaktiv produziert wurden, da die spätere Auswertung durch einen Forscher nicht intendiert war. Es lassen sich dabei prinzipiell zwei Kategorien von Spuren unterscheiden: *Abnutzung (Erosion)* und *Ablagerung (Akkumulation)*. *Erosionsmaße* geben Hinweise darüber, ob bestimmte Objekte besonders häufig oder besonders wenig benutzt wurden. Aus dem Abnutzungsgrad lässt sich dann beispielsweise ableiten, welche Relevanz der Gegenstand besaß bzw. ob ein besonderes Interesse an dem Gegenstand vorlag. Oft liegt allerdings eine Vielzahl von Erklärungsmöglichkeiten für Abnutzungserscheinungen vor, sodass sie nicht zuletzt deshalb in ihren Einsatzmöglichkeiten limitiert sind bzw. die Abnutzungsmaße vornehmlich zur Bestätigung anderer Messungen herangezogen werden. *Akkumulationsmaße* versuchen anhand der Ansammlung von Überresten oder Rückständen an Gegenständen u.a. Rückschlüsse auf kulturelle Eigenarten oder das Konsumverhalten der Probanden zu ziehen. Dabei können nahezu alle Arten von menschlichen Ablagerungen als Datenmaterial herangezogen werden. (Albrecht (1972), S. 245ff; Berg (2009), S. 287ff; Bungard/Lück (1995), S. 199; Mayrhofer (1993), S. 16f; Petermann/Noack (1999), S. 446f; Webb/Campbell/Schwartz/Sechrest/Grove (1981), S. 4ff)

Die Auswertung von Archivmaterial bzw. Dokumenten beschreibt eine wissenschaftliche Sekundäranalyse von Aufzeichnungen, die von den Verfassern der jeweiligen Dokumente nicht vorhersehbar war. Die Analyse von Dokumenten ist dabei allerdings nur dann als nicht-reaktiv anzusehen, wenn Datenerhebung und Datenauswertung deutlich voneinander getrennt sind und wenn Analysezweck und ursprünglich intendierter Verwendungszweck der Dokumente inkongruent sind. Verzerrungen der Validität von Forschungsergebnissen können sich darüber hinaus aus der Selektivität der Anlage von Aufzeichnungen sowie der Selektivität des Überlebens von Aufzeichnungen ergeben. So lässt sich der Inhalt bestimmter Texte nicht unbedingt auf den Forschungsgegenstand generalisieren oder es ergeben sich je nach Art und Verfasser eines Dokumentes unterschiedliche Aufbewahrungschancen. (Albrecht (1972), S. 250ff; Berg (2009), S. 271; Bungard/Lück (1995), S. 199f; Diekmann (2011), S. 653f; Mayrhofer (1993), S. 19f; Petermann/Noack (1999), S. 449f, S. 452f)

Prinzipiell lassen sich öffentliche und private Aufzeichnungen differenzieren, wobei die Hauptunterschiede in der Zugänglichkeit und Kontinuität des Datenmaterials liegt. Öffentliche Archive zeichnen sich in der Regel durch eine weitgehende Standardisierung und systematische Archivierung des Datenmaterials aus, sodass es bereits für wissenschaftliche Untersuchungen präpariert ist. Fortlaufende Aufzeichnungen erlauben dabei sogar oft auch Längsschnittuntersuchungen. BERG unterscheidet in Anlehnung an WEBB/CAMPBELL/SCHWARTZ/SECHREST/GROVE drei Kategorien öffentlicher Aufzeichnungen: (1) *commercial media accounts* beschreiben Datenmaterial, das sich aus Massenmedien ergibt (z.B. Zeitungen, Bücher, Zeitschriften, Fernsehmitschnitte, Videobänder, Comics, Landkarten etc.); (2) *actuarial records* beinhalten Datenmaterial, das vornehmlich für ein eingeschränktes

Publikum erstellt wurde, unter bestimmten Bedingungen allerdings auch von der Öffentlichkeit eingesehen werden kann (z.B. Geburts-, Sterbe-, Heirats- und Scheidungsurkunden, Versicherungs- und Bankaufzeichnungen, Besitzurkunden, demografische Aufzeichnungen etc.); (3) *official documentary records* betreffen Aufzeichnungen, Akten und den Nachrichtenverkehr von Schulen, Behörden, Krankenhäusern, Einzelhandelsunternehmen bzw. allen Organisationen, die solche Dokumente erstellen und aufbewahren. In der Regel sind solche Unterlagen nur für den internen Gebrauch bestimmt, finden manchmal aber auch ihren Weg in die Öffentlichkeit. Die Aufzeichnungen können dabei eine Vielzahl von Formen annehmen (z.B. Gerichtsmitschriften, Polizeiberichte, Zensusinformationen, Finanzbelege, Kriminalstatistiken, politische Reden, Regierungsakte, Parlamentsprotokolle, Schulaufzeichnungen, Frachtbriefe, Verkaufslisten, Memos, Emails, Mitschnitte von Meetings, Newsletter etc.) (Berg (2009), S. 272ff; Webb/Campbell/Schwartz/Sechrest/Grove (1981), S. 78ff)

Private Aufzeichnungen sind in der Regel für ein spezielles Publikum vorgesehen, weniger standardisiert und oft nur auf einen begrenzten Zeitabschnitt bezogen. In manchen Fällen stellen Dokumente dieser Art Fundsachen dar, die erst durch einen Forscher entdeckt werden und nicht im Hinblick auf spätere Auswertung erstellt wurden. In anderen Fällen werden die entsprechenden Dokumente von einem Forscher angefordert und speziell für diese Anfrage erstellt. Private Archive umfassen dabei im Wesentlichen Autobiografien, Memoiren, Tagebücher, Briefe, Heimvideos, Kunstgegenstände etc. und werden oft für Fallstudien herangezogen, da sie Selbstoffenbarungen und individuelle Definitionen historisch relevanter Situationen enthalten können. Wichtigstes Element privater Aufzeichnungen sind nach BERG Autobiografien, die Aussagen über die sozialen Rahmenbedingungen einer bestimmten Epoche, vorherrschende bzw. konkurrierende Ideologien von Gruppen machen oder Selbstreflexionen enthalten. Sie lassen sich in drei Kategorien differenzieren: (1) *comprehensive autobiographies* enthalten die Memoiren einer spezifischen Person und spannen einen Bogen über das gesamte Leben dieser Person; (2) *topical autobiographies* enthalten Ausschnitte aus dem Leben einer Person; (3) *edited autobiographies* stellen redaktionell bearbeitete Schriften dar, die bestimmte Aspekte des Lebens einer Person hervorheben oder verkürzen. (Berg (2009), S. 282ff; Webb/Campbell/Schwartz/Sechrest/Grove (1981), S. 162ff)

Kennzeichen nicht-reaktiver Beobachtungen sind, dass die Observation in der natürlichen Umgebung des Forschungsgegenstandes stattfindet, die beobachteten Subjekte sich der Observierung nicht bewusst sind, die Verhaltensbeobachtung keinen Eingriff in die natürlichen Abläufe des Feldes vollzieht und dass nach Möglichkeit kein direkter Kontakt zwischen Beobachter und untersuchten Personen besteht bzw. zustande kommt. Diese Kriterien werden im Grunde ausschließlich von der verdeckten und nicht-teilnehmenden Beobachtung erfüllt. Gegenstand solcher Beobachtungen können beispielsweise äußere Körpermerkmale, Kleidung, Mimik, rollenspezifische Gewohnheiten/Verhaltensweisen, Sprache etc. sein. (Mayrhofer (1993), S. 17f; Petermann/Noack (1999), S. 449f, S. 447f; Webb/Campbell/Schwartz/Sechrest/Grove (1981), S. 197ff)

Feldexperimente stellen Experimente dar, die in der natürlichen Umgebung des zu erforschenden Gegenstandes stattfinden, wobei alltägliche Ereignisse als experimentelles *treatment* verwendet werden. Typische Beispiele für nicht-reaktive Feldexperimente sind die *lost-letter-technique*, bei der adressierte und frankierte Briefe an öffentlichen Orten abgelegt werden und die Zahl der bei den Adressaten angekommenen Briefe als Indikator für die Einstellung gegenüber den Adressaten gewertet werden, und die *wrong-number-technique*, bei der ausgesuchte Personen von vermeintlichen Ausländern oder Einheimischen angerufen

werden, die unter einem Vorwand um eine Hilfeleistung bitten, und die Reaktion als Hinweis auf die Einstellung gegenüber bestimmten Bevölkerungsgruppen gewertet wird. (Bungard/Lück (1995), S. 200; Diekmann (2011), S. 630ff; Mayrhofer (1993), S. 21f; Petermann/Noack (1999), S. 451f) Zu den Vertretern dieser Gattung von Experimenten lassen sich mit Einschränkungen mitunter auch die auf GARFINKEL zurückgehenden Krisenexperimente (*breaching experiments*) zählen. In deren Rahmen die als unhinterfragt gültig angenommenen Regeln des Alltagshandelns durch die Unterwanderung dieses *common sense* und die daraus resultierenden Verhaltensweisen offenlegt werden sollen. (Garfinkel (1963); Garfinkel (2010))

Als Prototyp für den Modus und das Instrumentarium eines nicht-reaktiven Experimentes lässt sich möglicherweise das qualitative Experiment heranziehen. Gegenstand qualitativer Experimente sind soziale Phänomene im Allgemeinen, wobei der Eingriff in den Gegenstand innerhalb dessen natürlicher Umgebung stattfindet und eine Variation der Struktur des Gegenstandes beinhaltet. Ziel ist es dabei, die soziale Struktur des Gegenstandes offenzulegen, wobei die Struktur das Typische, Spezifische und Besondere des Gegenstandes ist. Es sollen insbesondere die Verhältnisse, Relationen und Abhängigkeiten mit bzw. zu anderen Objekten offengelegt werden. Bei der Variation der Struktur werden verschiedene Handlungsstrategien angewendet, die sich mit Minimierung/Maximierung, Testen der Grenzen und Adaption beschreiben lassen. Konkret kann durch verschiedene Möglichkeiten der *Gliederung* des Gegenstandes (*Separation/Segmentation* bzw. *Kombination*) eine Teilung bzw. Gliederung eines Gegenstandes herbeigeführt werden, um unterschiedliche Arten von Sinn zu erzeugen oder die Teile des Gegenstandes auf andere Art und Weise zusammensetzen als zuvor. Möglich sind auch verschiedene Methoden zur *Einschränkung* bzw. *Ausdehnung* des Gegenstandes (*Reduktion/Abschwächung* bzw. *Adjektion/Intensivierung*), in deren Rahmen entweder Teile des Gegenstandes entfernt oder hinzugefügt bzw. verstärkt werden und beobachtet wird, ob und wie sich Veränderungen manifestieren. Zuletzt kann der Gegenstand auch Formen einer *Umwandlung* (*Substitution* bzw. *Transformation*) erfahren, in deren Folge einzelne Teile des Gegenstandes durch neue ersetzt werden oder eine Umwandlung des Gegenstandes in einen anderen stattfindet, wobei noch Merkmale des ursprünglichen Gegenstandes beibehalten werden. (Kleining (1986); Lamnek (2010), S. 582ff)

Wie bereits beschrieben, liegt der große Vorteil nicht-reaktiver Verfahren in der Eliminierung von Verzerrungen, die aus der Interaktion von Forscher und Subjekt und dem Rollenverhalten des Subjektes resultieren. Ersteres wird dadurch herbeigeführt, dass der Forscher als solcher nicht erkennbar ist bzw. allenfalls indirekter Kontakt zwischen Forscher und Subjekt besteht. Letzteres wird durch Anwendung lebensnaher Indikatoren innerhalb der vertrauten Lebenswelt der Subjekte erreicht. Daneben sind mit nicht-reaktiven Verfahren allerdings auch eine Reihe von Nachteilen verbunden (Bungard/Lück (1995), S. 200f; Diekmann (2011), S. 655f; Mayrhofer (1993), S. 24f; Petermann/Noack (1999), S. 444, S. 456ff; Webb/Campbell/Schwartz/Sechrest/Grove (1981), S. 275ff):

- (1) *Stichprobenprobleme/-auswahl*: Aufgrund der Datenerhebung in natürlicher Umgebung und der vergleichweisen Passivität des Erhebungsvorgangs unterliegen nicht-reaktive Verfahren Einschränkungen, was die Repräsentativität des Samples angeht. Zusätzlich lässt sich die Grundgesamtheit der Fälle kaum bestimmen.
- (2) *Prüfung der Gütekriterien*: Hinsichtlich der Existenz von Objektivität, Reliabilität und Validität ergeben sich abhängig vom jeweiligen nicht-reaktiven Verfahren teilweise erhebliche Unterschiede. Im Rahmen von Beobachtungen lässt sich oft keine Objektivität herstellen, da der Einsatz mehrerer parallel aktiver Beobachter zu auffällig ist und

Veränderungen im Forschungsgegenstand nach sich ziehen kann. Hieraus ergeben sich wiederum Konsequenzen für die Reliabilität, die an die Güte des verwendeten Beobachtungssystems und die Fähigkeiten der eingesetzten Beobachter gebunden ist. Reliabilitätsprobleme, insbesondere hinsichtlich der Retest-Reliabilität, ergeben sich auch für nicht-reaktive Experimente. Hier können infolge von Mehrfachmessungen an den gleichen Probanden Verhaltensveränderungen auftreten. Die Analyse physikalischer Spuren und inhaltsanalytischer Untersuchung von Dokumenten sind dagegen weitestgehend frei von Problemen, welche die Objektivität oder Reliabilität betreffen. Für alle nicht-reaktiven Verfahren gilt dagegen, dass Schwierigkeiten bezüglich der Konstruktvalidität auftreten können. So können bei der Validierung von Messergebnissen durch die parallele Anwendung anderer Verfahren Abweichungen zwischen den Resultaten bzw. widerstreitende Ergebnisse auftreten. Es stellt sich dann die Frage welches Ergebnis man als valide annimmt.

- (3) *Einfluss von Versuchsleiter-Effekten*: Verzerrungen können durch spezifische Erwartungen des Versuchsleiters hinsichtlich der Resultate verzerrt werden.
- (4) *Absicherung der inhaltlichen Schlussfolgerungen*: Es muss nicht zwingend eine inhaltliche Verbindung zwischen interessierender Variable und gewähltem Indikator bestehen. Möglicherweise liegt kein oder nur ein indirekter Zusammenhang vor, der durch Störgrößen oder intervenierende Variablen vermittelt wird. Zusätzlich ist die Erfassung von latenten Variablen, wie Handlungsmotiven, Emotionen oder ein Verhalten bedingende kognitive Prozesse, mit Hilfe nicht-reaktiver Verfahren nicht ohne Weiteres umzusetzen.
- (5) *Ethische Probleme*: Eine systematische Erhebung von Daten, ohne dass die betroffenen Personen davon Kenntnis besitzen, ist ethisch zumindest fragwürdig. Gleiches gilt dafür, Personen aus purem Erkenntnisinteresse potentiell gefährlichen Situationen auszusetzen.

2.2.2.3 Implizite Messverfahren bzw. Maße

Implizite Messverfahren bilden das Pendant zu den bereits beschriebenen expliziten Messverfahren und setzen im Gegensatz zu diesen keine bewusste Aufmerksamkeit seitens der Befragten für das interessierende Konstrukt voraus. (Jonas/Stroebe/Hewstone (2007), S. 207)

„The term *implicit* has come to be applied to measurement methods that avoid requiring introspective access, decrease the mental control available to produce the response, reduce the role of conscious intention, and reduce the role of self-reflective, deliberative processes.“ (Nosek/Greenwald/Banaji (2007), S. 267)

Während explizite Messmethoden dazu dienen, Einstellungen und Kognitionen zu erfassen, zu denen die Probanden einen bewussten Zugang besitzen bzw. die einer willentlichen Artikulation zugänglich sind, lassen sich mit Hilfe impliziter Messmethoden spontane und unbewusste Assoziationen bzw. Einstellungen, die keiner bewussten Kontrolle unterliegen, bestimmen. (Asendorpf/Banse/Mücke (2002), S. 381; Brunel/Tietje/Greenwald (2004), S. 386f; Gawronski/De Houwer (2014), S. 283; Hefner (2013), S. 250ff; Rauthmann (2017), S. 208f) Theoretische Grundlage impliziter Methoden ist die Annahme, dass sich innerhalb impliziter Kognitionen vergangene Erfahrungen manifestieren, die wiederum die aktuelle Verhaltensperformanz beeinflussen. Diese Erfahrungen müssen dabei keineswegs bewusst zugänglich sein, um einen Effekt auf das Handeln oder die Einstellungen einer Person zu haben.

2. Theorie: Methodik

(Fazio/Olson (2003), S. 302; Greenwald/Banaji (1995), S. 4f; Nosek/Greenwald/Banaji (2007), S. 266)

„An implicit **C** is the introspectively unidentified (or inaccurately identified) trace of past experience that mediates **R**. In this template, **C** is the label for a construct (such as attitude), and **R** names the category of responses (such as object-evaluative judgements) assumed to be influenced by that construct.“ (Greenwald/Banaji (1995), S. 5)

Solche impliziten Kognitionen können dabei impliziten Einstellungen, einem impliziten Selbstwertgefühl oder impliziten Stereotypen entsprechen. Sie wirken sich in der Folge auf Gefühle, Gedanken und Handlungen gegenüber anderen sozialen Objekten aus, bestimmen die Bewertung selbst-assoziiierter und selbst-dissoziierter Objekte und schreiben Mitgliedern bestimmter sozialer Gruppen Attribute und Fähigkeiten zu. (Greenwald/Banaji (1995), S. 8, S. 11, S. 15)

HEFNER beschreibt dies in Anlehnung an ANDERSON (Anderson (1983)) und GREENWALD/BANAJI (Greenwald/Banaji (1995)) als Aktivierung einer assoziativen Netzwerk-Struktur:

„Werden Elemente des Netzwerkes aktiviert, so breitet sich die Aktivierung auf andere Elemente aus, die in einer inhaltlichen Verbindung zu den aktivierten stehen. Die Aktivierung eines mentalen Konzeptes zieht also die Aktivierung anderer, assoziierter Konzepte nach sich.“ (Hefner (2013), S. 252)

Die beschriebenen Assoziationen unterscheiden sich in ihrer Stärke voneinander, wobei starke Assoziationen schneller aktiviert werden, während schwache Assoziationen eine höhere Verzögerung aufweisen. Ziel impliziter Messverfahren ist es zu erfassen, welche kognitiven Konzepte besonders eng miteinander verknüpft sind. Dabei bedienen sie sich Antwortlatenzzeiten bzw. Reaktionszeiten und Fehlerraten bei der Bearbeitung der jeweiligen Testaufgaben als Indikatoren. Um zu verhindern, dass die Messergebnisse durch höhere kognitive Prozesse beeinflusst werden und zu gewährleisten, dass vornehmlich bzw. ausschließlich spontane und automatische Reaktionen für das Messergebnis verantwortlich sind, wird über Zeitdruck während der Aufgabenbearbeitung eine Verminderung der Kontrolle der Probanden über ihr Antwortverhalten realisiert. (Gawronski/De Houwer (2014), S. 298f; Hefner (2013), S. 251f; Krohne/Hock (2015), S. 335)

Es wird angenommen, dass implizite Messverfahren im Gegensatz zu expliziten Messverfahren dazu in der Lage sind, soziale Reaktivitätseffekte, wie Verfälschungstendenzen bzw. sozialerwünschtes Antwortverhalten zu verhindern. Neben der Reduktion der Kontrolle über die Messresultate zeichnet hierfür auch die Tatsache verantwortlich, dass die Probanden meist nicht bemerken, dass mit den Testaufgaben Einstellungen, Stereotype o.ä. erfasst werden. (Fazio/Olson (2003), S. 300, S. 302f; Gawronski/De Houwer (2014), S. 298f; Hefner (2013), S. 251) Zwar fällt der Effekt intentionaler Verfälschung auf die Ergebnisse impliziter Messverfahren niedriger aus als für Selbstbeurteilungsmaße, allerdings können auch implizite Messmethoden nicht als verfälschungssicher angenommen werden. Vielmehr zeigt sich, dass, im Sinne eines Trainingseffektes, Erfahrungen mit bestimmten impliziten Verfahren die Vortäuschung einer bestimmten Einstellung ermöglichen. Außerdem scheint durch Verlangsamung der Reaktionszeiten oder eine Erhöhung der Fehleranzahl die prinzipielle Möglichkeit für „*faking bad*“ gegeben zu sein. Es zeigte sich zudem, dass Probanden dazu in der Lage sind, infolge von *faking*-Instruktionen erfolgreich die Ergebnisse von IATs zu verfälschen. Ebenso spielen Kontexteffekte bezüglich der im Rahmen der Testaufgaben verwendeten Stimuli

2. Theorie: Methodik

eine Rolle. Auch dürfen Abweichungen im Messergebnis von expliziten und impliziten Messverfahren nicht per se als Ausdruck eines Bias interpretiert werden. Solche Unterschiede sind oft inhaltlichen Diskrepanzen der Form geschuldet, dass die erfassten Konstrukte nicht vollständig kongruent sind bzw. verschiedene Aspekte eines Konstruktes gemessen werden. Daneben spielen auch die strukturellen Unterschiede in den Aufgabenstellungen von expliziten und impliziten Messverfahren eine Rolle. (De Houwer (2006), S. 16ff, S. 21ff; Fazio/Olson (2003), S. 317f; Gawronski/De Houwer (2014), S. 298ff; Goodall (2011), S. 214f)

DE HOUWER begreift implizite Maße vornehmlich als Messergebnis und weniger als Messprozedur bzw. Messverfahren. Implizite Maße sind dabei im Wesentlichen durch drei funktionelle Eigenschaften definiert: So dienen implizite Maße als Indikator für Einstellungen und Kognitionen, wobei (1) die Probanden sich gar nicht bewusst sein müssen, dass ein bestimmtes Konstrukt gemessen wird, (2) die Probanden keinen bewussten Zugang zu dem jeweiligen interessierenden Konstrukt besitzen müssen oder (3) die Probanden das Messergebnis nicht willentlich beeinflussen können bzw. die Ausprägung des Konstrukts nicht bewusst kontrollieren können. (Asendorpf/Banse/Mücke (2002), S. 381; Brunel/Tietje/Greenwald (2004), S. 386ff; De Houwer (2006), S. 12f; De Houwer/Moors (2007), S. 180f; Fazio/Olson (2003), S. 300, S. 302f) Anders ausgedrückt:

„(...) the term ‚implicit measure‘ refers to certain functional properties of measurement outcomes: The outcome functions as an index of an attitude or cognition despite the fact that participants are unaware of the impact of the attitude or cognition on the outcome, are not aware of the attitude or outcome, or have no control over the outcome.“ (De Houwer (2006), S. 12)

Dabei muss bestimmt werden, hinsichtlich welcher der hier genannten drei funktionalen Eigenschaften es sich um ein implizites Maß handelt, wobei nicht jedes implizite Maß über jede der funktionalen Eigenschaften verfügen muss bzw. die funktionalen Eigenschaften nicht notwendigerweise gemeinsam auftreten müssen. Die Spezifizierung eines Maßes als implizit kann dabei ausschließlich auf empirischem Wege erfolgen, sodass jede der Eigenschaften einzeln auf Gültigkeit hin überprüft werden muss. (De Houwer (2006), S. 13f; De Houwer/Moors (2007), S. 180ff)

Nach DE HOUWER weisen implizite Maße erhebliche Parallelen zu sogenannten automatischen Maßen auf, sodass „one can define an implicit measure as a measurement outcome that reflects a certain attitude or cognition in an automatic manner, where ‚automatic‘ needs to be specified in terms of the presence of one or more functional features“ (De Houwer (2006), S. 14). Um zu bestimmen, in welcher Art und Weise die Merkmale eines Maßes ausgestaltet sein müssen, um den funktionalen Eigenschaften eines impliziten Maßes zu entsprechen, nutzt DE HOUWER daher die Charakteristika automatischer Maße. (Houwer/Moors (2007), S. 182ff) und definiert implizite Maße wie folgt:

„(...) measurement outcomes that reflect the to-be-measured construct by virtue of processes that are uncontrolled, unintentional, goal-independent, purely-stimulus-driven, autonomous, unconscious, efficient, or fast.“ (De Houwer/Moors (2007), S. 188f)

Implizite Messmethoden erfassen individuelle Unterschiede vornehmlich anhand der Berechnung von Reaktionszeitdifferenzen. Das Instrumentarium umfasst dabei mittlerweile eine Reihe von Verfahren, die sich im Wesentlichen als Priming- und Kategorisierungsaufgaben beschreiben lassen. Daneben werden aber auch Wortergänzungs- und

2. Theorie: Methodik

Satzvervollständigungsaufgaben sowie physiologische Messungen den impliziten Methoden zugerechnet. Zum erweiterten Kreis impliziter Messverfahren werden oft auch projektive Verfahren und Objektive Tests gezählt. (De Houwer (2006), S. 11; Fazio/Olson (2003), S. 298ff; Schmukle/Egloff (2011), S. 73ff)

Allgemein lässt sich Priming bzw. dessen Ergebnis wie folgt beschreiben:

„Primingeffekte liegen vor, wenn die Konfrontation mit einem Reiz A die Verarbeitung eines gleichzeitig oder nachfolgend dargebotenen Reizes B beeinflusst.“ (Krohne/Hock (2015), S. 339)

Beim *affektiven* bzw. *evaluativen Priming* werden die Probanden zunächst kurz einem Einstellungsobjekt (Prime) ausgesetzt und bekommen im direkten Anschluss positive oder negative Worte (Targets) präsentiert. Die Probanden sollen nun möglichst schnell die Targets als positiv oder negativ kategorisieren. Je nachdem, ob der Prime für die Probanden positiv oder negativ assoziiert ist, fällt die Reaktionszeit auf ein positives oder negatives Target kürzer aus. Auf diese Weise kann indirekt festgestellt werden, welche Einstellung die Probanden gegenüber dem Einstellungsobjekt haben. Prinzipiell lassen sich mit Hilfe dieser Methode Evaluationen bezüglich jeglichen Objekts erfassen, das sich als Prime darstellen lässt. Eine Abwandlung dieser Basisvariante des Primings sind die *Semantic Priming Tasks*, bei denen die semantische Beziehung der Stimuli anstatt deren evaluativer Bedeutung im Vordergrund steht. (Gawronski/De Houwer (2014), S. 286ff; Jonas/Stroebe/Hewstone (2007), S. 210f; Schmukle/Egloff (2011), S. 91ff)

Der *Implicit Association Test (IAT)* baut auf dem Prinzip des Primings auf und stellt gleichzeitig das prominenteste Beispiel impliziter Messverfahren dar. Der IAT beinhaltet zwei binäre Kategorisierungsaufgaben, deren Stimuli während der Bearbeitung sowohl in Richtung der Präferenz der Probanden als auch dieser Präferenz entgegengesetzt kombiniert werden. Die hintergründige Idee ist, dass bei Reizassoziationen, die der Präferenz der Probanden entsprechen schneller reagiert wird und die Antworten weniger Fehler enthalten, als wenn die Reizassoziationen nicht der Präferenz der Probanden entsprechen. (Gawronski/De Houwer (2014), S. 284f; Krohne/Hock (2015), S. 338f; Schmukle/Egloff (2011), S. 93ff)

Ein typischer IAT setzt sich dabei aus insgesamt fünf Blöcken zusammen. Drei dieser Blöcke stellen Trainingsphasen dar, während die verbleibenden zwei Blöcke zur Kalkulation des sogenannten *IAT-Effektes*³ herangezogen werden. In jedem Block soll die Kategorisierung der Stimuli möglichst schnell erfolgen. Im ersten Block (*initial target-concept discrimination*) wird das interessierende Zielkonzept eingeführt, wobei die gezeigten Stimuli per Tastendruck einer von zwei Kategorien zugeordnet werden sollen (z.B. alt vs. jung, schwarz vs. weiß, dick vs. dünn etc.). Im zweiten Block (*initial attribute discrimination*) werden den Probanden Bewertungsattribute präsentiert, die als positiv oder negativ klassifiziert werden sollen (z.B. Vergnügen, Freude, Friede, Qual). Die Zuordnung erfolgt dabei wieder über die gleichen Tasten wie im ersten Block, wobei jeweils die als „gut“ bzw. „schlecht“ eingeordneten Begriffe mit einer der Kategorien aus dem ersten Block verknüpft werden. Im dritten Block (*initial combined task*) werden beide Kategorisierungsaufgaben kombiniert, gleichzeitig stellt er die erste Messphase dar. Die Probanden bekommen hier abwechselnd sowohl Aspekte des Zielkonzeptes als auch positiv und negativ konnotierte Begriffe präsentiert und sollen diese entsprechend der Trainingsphase per Tastendruck kategorisieren. Die Zuordnung soll dabei so schnell und so

³ Differenz zwischen den Reaktionszeiten in Block 3 und Block 5 (Fischer/Asal/Krueger (2014), S. 105; Jonas/Stroebe/Hewstone (2007), S. 211f; Rauthmann (2017), S. 210)

fehlerfrei wie möglich erfolgen. Der vierte Block (*reversed target-concept discrimination*) beinhaltet wieder eine Trainingsphase, innerhalb derer die Probanden auf eine Umkehrung der im ersten Block erlernten Tastenzuordnung konditioniert werden. War zuvor die eine Kategorie der Zielkonzepte mit positiven Attributen und die andere Kategorie der Zielkonzepte mit negativen Attributen verknüpft, erfolgt nun eine Umpolung. Im fünften Block (*reversed combined task*) bzw. der zweiten Messphase werden den Probanden wieder abwechselnd sowohl die Zielkonzepte als auch die Attribute präsentiert, wobei die Kategorisierung entsprechend der im vierten Block erlernten Assoziation vollzogen werden soll. In der Folge wird ermittelt, ob sich die Reaktionszeiten zwischen dem dritten und dem fünften Block unterscheiden. Dabei sollten sich aus Irritationen resultierende Reaktionsverzögerungen ergeben, wenn die Probanden bei der Zuordnung mit einer ihrer Ansicht nach inkompatiblen Kombination von Zielkonzept und Bewertungsattribut konfrontiert sind. Werden die präsentierten Kombinationen von den Probanden dagegen als kompatibel angenommen, sollte die Zuordnung leichter fallen und die Reaktionszeit entsprechend kürzer bemessen sein. Aus der Differenz der Reaktionszeiten kann dann die Einstellung zu den Zielkonzepten abgeleitet werden. (Gawronski/De Houwer (2014), S. 285; Jonas/Stroebe/Hewstone (2007), S. 211f; Krohne/Hock (2015), S. 339f; Schmukle/Egloff (2011), S. 93ff)

Das Konzept des IAT stellt die Basis für eine Vielzahl weiterer impliziter Messverfahren dar. Diese sind entweder durch dessen strukturelle Verfasstheit inspiriert oder stellen Weiterentwicklungen des ursprünglichen Konzeptes dar: *Go/No-Go Association Task (GNAT)*, *Extrinsic Affective Simon Task (EAST)*, *Sorting Paired Features Task (SPF)*, *Implicit Relational Assessment Procedure (IRAP)*, *Single Category IAT (SC-IAT)*. (Fazio/Olson (2003), S. 300; Gawronski/De Houwer (2014), S. 289ff; Goodall (2011), S. 215f; Schmukle/Egloff (2011), S. 104ff)

Mit der *Affect Misattribution Procedure (AMP)* wurde ein Verfahren entwickelt, das versucht, verschiedene Probleme des IAT und von Priming-Maßen aufzufangen. Bezüglich des IAT stellt dies u.a. ein mangelhaftes Verständnis für die den Messergebnissen zugrundeliegenden Mechanismen dar, während *evaluatives Priming* Reliabilitätsprobleme aufweist. In der Basisversion der AMP wird den Probanden zunächst kurz ein Prime gezeigt, woran sich die kurze Präsentation eines chinesischen Schriftzeichens anschließt. Dieses Schriftzeichen wird in der Folge durch ein Schwarz-Weiß-Muster ersetzt und die Probanden sollen dann angeben, ob sie das Schriftzeichen verglichen mit einem durchschnittlichen chinesischen Schriftzeichen als mehr oder weniger ansprechend empfunden haben. Es zeigt sich, dass das Schriftzeichen dann positiver bewertet wird, wenn der Prime positiv besetzt ist und umgekehrt. Die AMP basiert dabei auf der Annahme, „that when conditions are ambiguous, participants will misattribute their affective responses from one source (i.e., a prime) to another source (i.e., a Chinese pictograph)“ (Goodall (2011), S. 218). Anders als bei anderen impliziten Messverfahren basiert der Messvorgang der AMP nicht auf der Erfassung von Reaktionszeiten. (Gawronski/De Houwer (2014), S. 288f; Goodall (2011), S. 218)

KROHNE/HOCK unterscheiden implizite Messverfahren danach, ob sie sich nur auf bestimmte Merkmale anwenden lassen bzw. nur eng umgrenzte Eigenschaften erfassen können oder ob sie in der Lage sind, globalere Eigenschaften zu messen bzw. auf verschiedene Konstrukte angewendet werden können. Während der IAT zur letztgenannten Gruppen gehört, stellen der *Emotional Stroop-Task* und die *Dot-Probe-Task* Vertreter der erstgenannten Gruppe impliziter Methoden dar und zielen auf die Bestimmung der Ängstlichkeit von Probanden. (Krohne/Hock (2015), S. 335)

Bei der *Emotional Stroop-Task* werden den Probanden Worte mit Bedrohungscharakter und neutrale Worte präsentiert, die in verschiedenen Farben geschrieben sind bzw. mit verschiedenen Farben hinterlegt sind. Aufgabe der Probanden ist es, die Farbe der Stimuli möglichst schnell zu bestimmen. Es wird angenommen, dass bei ängstlichen Personen die Farbbenennungszeit für bedrohliche Worte länger ausfällt als für neutrale Worte. (Krohne/Hock (2015), S. 335f; Schmukle/Egloff (2011), S. 88f) Dies wird damit begründet, „dass bei ängstlichen Personen die affektive Komponente der zumindest partiell verarbeiteten Reizwörter mit der Farbbenennungsaufgabe stärker interferieren soll als bei Niedrigängstlichen“ (Schmukle/Egloff (2011), S. 88).

Im Rahmen der *Dot-Probe-Task* werden den Probanden vertikal angeordnete Wortepaare präsentiert, die sich jeweils aus einem bedrohungsassoziierten und einem affektiv neutralen Wort zusammensetzen. Die Worte erscheinen dabei je nach Kategorie abwechselnd an oberer oder unterer Stelle, wobei gelegentlich entweder der bedrohliche oder der neutrale Stimulus durch einen Punkt ersetzt wird. Die Aufgabe der Probanden besteht darin möglichst schnell eine Taste zu drücken, wenn der Punkt erscheint. (Krohne/Hock (2015), S. 336ff; Schmukle/Egloff (2011), S. 89ff) Es wird davon ausgegangen, „dass die Reaktionszeiten auf den Punkt dann relativ schnell ausfallen, wenn die visuelle Aufmerksamkeit auf demjenigen Stimulus lag, der in der Position dargeboten wurde, die der des Punktes entspricht“ (Schmukle/Egloff (2011), S. 89). Ängstliche Personen sollten demnach schneller auf das Ersetzen eines bedrohlichen Wortes reagieren als auf das Ersetzen eines neutralen Wortes.

Was die Reliabilität impliziter Verfahren angeht, so zeichnet sich ein vergleichsweise heterogenes Bild. Während *evaluatives Priming* eher niedrige Retest-Reliabilitätswerte aufweist, reicht die Retest-Reliabilität des *IAT* beispielsweise annähernd an die von Selbstbeurteilungsmaßen heran. Hinsichtlich der internen Konsistenz liegen solche auf der Erfassung von Latenzzeiten beruhenden impliziten Messverfahren allerdings vor anderen impliziten Maßen. Die Tatsache, dass für eine Vielzahl impliziter Messverfahren nur eine geringe Retest-Reliabilität zu finden ist, wird dabei mit der Instabilität impliziter Kognitionen erklärt. (Fazio/Olson (2003), S. 311f; Gawronski/De Houwer (2014), S. 287; S. 293, S. 298; Goodall (2011), S. 208ff; Robinson/Neighbors (2006), S. 123; Schmukle/Egloff (2011), S. 96f)

Die Validität von Priming-Aufgaben ist im Wesentlichen an die Repräsentativität der verwendeten Stimuli gebunden. Im Allgemeinen gilt, dass die Ergebnisse impliziter Messverfahren meist keine konvergente Validität aufweisen, die Korrelationen mit alternativen Messungen der gleichen Zieldimension fallen somit eher gering aus. Die Gründe hierfür liegen womöglich darin, dass die miteinander in Beziehung gesetzten Maße zwar das gleiche Konstrukt erfassen, dabei aber unterschiedliche kognitive Prozesse berühren. Die Kriteriumsvalidität impliziter Messverfahren fällt vergleichsweise hoch aus. In diesem Zusammenhang ist insbesondere die prädiktive Validität hervorzuheben, sodass implizite Messverfahren durchaus zur Vorhersage von Verhalten herangezogen werden können. (Fazio/Olson (2003), S. 305ff; Gawronski/De Houwer (2014), S. 294; Goodall (2011), S. 208, S. 211; Robinson/Neighbors (2006), S. 123f; Schmukle/Egloff (2011), S. 98ff)

2.2.2.4 Projektive Verfahren

Im Rahmen projektiver Verfahren bekommen die Probanden vergleichsweise gering strukturierte, mehrdeutige Testvorlagen (Gegenstand sind bspw. Wortassoziation, verbale

2. Theorie: Methodik

Ergänzung, Formdeutung, thematische Apperzeption, spielerische oder zeichnerische Gestaltung, Farbwahl bzw. Farbgestaltung sowie Bildauswahl) vorgelegt, die je nach Aufgabe interpretiert, rezipiert, vervollständigt oder angeordnet werden sollen. Dabei gehen projektive Verfahren davon aus, dass man aus den Äußerungen von Menschen (z.B. Erzählungen, Assoziationen, Zeichnungen etc.) Rückschlüsse auf deren Persönlichkeitsstruktur ziehen kann. Die persönlichen Äußerungen werden dabei als Wesensmerkmale der jeweiligen Person interpretiert. (Axhausen (1999), S. 461; Leichsenring/Hiller (1994), S. 162)

„Projektion bedeutet dabei im weiten Sinn, daß Wertvorstellungen und Wünsche genauso wie Motive, Bedürfnisse oder Triebimpulse auf Situationen, Menschen und Objekte außerhalb der eigenen Person übertragen werden.“ (Axhausen (1999), S. 461)

Die Uneindeutigkeit der verwendeten Stimuli ist dabei gleichermaßen Stärke wie Schwäche der projektiven Verfahren. So besteht die Schwierigkeit zunächst in einer exakten Deutung der Projektionen. Das heißt: Es lässt sich von den Projektionen auf bestehende Eigenschaften der Probanden schließen oder es handelt sich eher um Eigenschaften, die sich die Probanden für sich selbst wünschen bzw. es handelt sich um Eigenschaften, denen sie ablehnend gegenüberstehen. Mit anderen Worten: Bei den Ergebnissen projektiver Verfahren handelt es sich nicht unbedingt um präzise und objektiv überprüfbare Interpretationen. Gleichzeitig ermöglicht die Mehrdeutigkeit des Aufforderungscharakters einen vergleichsweise unverfälschten Ausdruck der Persönlichkeit der Probanden. Durch die spezielle Gestaltung des Testmaterials projektiver Verfahren wird die Wahrscheinlichkeit sozialerwünschten Antwortverhaltens gesenkt. Dies liegt nicht zuletzt daran, dass projektive Verfahren in der Regel eine Aktivierung unbewusster Erlebnisinhalte bewirken, die per definitionem keiner bewussten Kontrolle zugänglich sind und dementsprechend nicht willentlich beeinflusst werden können. (Axhausen (1999), S. 462f; Wittkowski (2011), S. 304)

Projektive Verfahren haben somit mit einer mangelnden Messgenauigkeit zu kämpfen, deren Konsequenzen sich auf verschiedenen Ebenen manifestieren (Axhausen (1999), S. 463):

- (1) *Testmaterial*: Unbestimmtheit des Aufforderungscharakters lässt exakte Bestimmung des Stimulus offen.
- (2) *Testsituation*: Potentielle Suggestionen seitens des Versuchsleiters.
- (3) *Projektion*: Theoretische Unschärfe des hierbei wirksamen Mechanismus.
- (4) *Auswertung*: Problem der Eindeutigkeit und mangelnde Vergleichbarkeit der Messergebnisse.

Es existiert eine Vielzahl von projektiven Verfahren, die jeweils unterschiedliche Erkenntnisziele haben. Das Erkenntnisinteresse reicht dabei von emotionalen Konflikten, der Abbildung der Persönlichkeitsstruktur, der aktuellen Lebenssituation, psychiatrische Diagnostik bis hin zu Erfassung affektiven Verhaltens. AXHAUSEN unterscheidet hierbei Wortassoziations- und verbale Ergänzungsaufgaben, Formdeutungsverfahren (z.B. Rohrschach-Test), thematische Apperzeptionsverfahren, spielerische und zeichnerische Gestaltungsaufgaben, Farbwahl- und Farbgestaltungsverfahren sowie Bildwahlverfahren. (Axhausen (1999), S. 466ff)

Nach WITTKOWSKI findet insbesondere im Bereich verbal-thematischer Verfahren, zeichnerischer und spielerischer Gestaltungsverfahren sowie semi-projektiver Verfahren eine Weiterentwicklung statt bzw. die genannten Verfahrensklassen werden verglichen mit anderen Verfahren häufiger in der Praxis angewendet. Im Rahmen verbal-thematischer Verfahren werden aus Geschichten, Deutungen oder Gesprächssequenzen der Probanden zu

gegenständlichen, allerdings mehrdeutigen Bildern Folgerungen hinsichtlich deren Persönlichkeitsdispositionen abgeleitet. Typischer Vertreter dieser Gruppe projektiver Verfahren ist der thematische Apperzeptionstest, dessen grundlegende Annahme darin besteht, dass in die Interpretation mehrdeutiger Situationen in Form einer Bildvorlage gegenwärtige und frühere Erlebnisse einfließen und auf diese Weise bedeutsame Aspekte der Persönlichkeit der Probanden zugänglich werden. Es wird weiterhin davon ausgegangen, dass die Probanden unbemerkt und unzensiert Inhalte mitteilen, indem sie ihr eigenes Selbst auf die Bildvorlage projizieren. (Wittkowski (2011), S. 300ff)

Zeichnerische Gestaltungsverfahren beinhalten eine Zeichenaufforderung an die Probanden, wobei die zeichnerischen Leistungen sowohl nach formalen als auch nach inhaltlichen Aspekten analysiert werden. Zu unterscheiden sind thematische Zeichentests (themenbezogenes Zeichnen ohne Vorlage; z.B. ein Mensch, ein Tier, ein Baum, die eigene Familie in Tiergestalt) und athematische Zeichentests (z.B. Vervollständigung von Zeichnungen, freies Zeichnen ohne Vorlage). Im Fokus des Interesses stehen dabei die Gründe der Probanden für den Gegenstand und die Art und Weise der Zeichnung bzw. die emotionale Relevanz der gezeichneten Inhalte. Dabei werden entweder der Prozess der Anfertigung der Zeichnung (Verhalten und sprachliche Äußerungen), das Produkt des Prozesses (Zeichnung) oder beides zu Auswertungszwecken betrachtet. (Wittkowski (2011), S. 354ff)

Spielerische Gestaltungsverfahren gehen davon aus, dass „im freien und spontanen Handeln die Art und Weise zum Ausdruck kommt, in der sich eine Person mit ihrer Umwelt auseinandersetzt“ (Wittkowski (2011), S. 360). Besondere Eigenart des Spiels ist dabei, dass es in einer durch das Spielmaterial geschaffenen Rahmung stattfindet. Gegenstand der Auswertung sind sowohl das Spielprodukt als auch der Spielverlauf bzw. die Spielhandlung. Das Spielprodukt wird dabei als „die Spur von etwas Eigentlichem, Unsichtbarem, Dahinterliegendem“ angesehen, sodass durch eine inhaltliche Interpretation Ursachen für psychische Probleme der Probanden aufgedeckt werden können. Ziel der Analyse des Spielverlaufs ist es, die kognitiven und kreativen Fähigkeiten der Probanden sowie ihre Reflektion der Umwelt zu erfassen. (Wittkowski (2011), S. 360f)

Semi-projektive Verfahren nehmen an, „dass Erleben und Verhalten sowohl von überdauernden Dispositionen als auch von Merkmalen der jeweiligen Situation bestimmt wird“ (Wittkowski (2011), S. 370). Der Schwerpunkt WITTKOWSKIs liegt dabei auf der sogenannten Gitter-Technik, die versucht, gleichermaßen Dispositionen wie situationsspezifische Einflüsse zu erfassen. Dispositionen entsprechen dabei impliziten Motiven, die als Basis für explizite Motive dienen, welche sich situationsabhängig ergeben. Letztere stellen Selbstbeschreibungen der Probanden dar, die sich als Reaktion auf äußere soziale Einflüsse einstellen und sich als Prognosegrundlage für die aktuellen Einstellungen, Werte und Ziele der Probanden anbieten. Die Gitter-Technik verwendet zwei unterschiedliche Untersuchungsmethoden, um die verschiedenen Motivebenen abzubilden. Die impliziten Motive der Probanden werden analog zum Apperzeptionstest mittels der Interpretation mehrdeutigen Bildmaterials erfasst, während die expliziten Motive anhand der Beantwortung von motivationsbezogenen Aussagen erhoben werden. (Wittkowski (2011), S. 370f)

2.2.2.5 Objektive Tests

2.2.2.5.1 Begriffsbestimmung

Der Begriff „Objektivität“ kann innerhalb der empirischen Sozialforschung verschiedene Bedeutungen annehmen: *Objektivität als Gütekriterium von Messungen, Objektivität im Sinne von Fairness, Objektivität in Form objektiver Merkmalsindikatoren und Objektivität im Sinne von Unverfälschbarkeit bzw. in Form sogenannter Objektiver Tests.*

Am bekanntesten ist hierbei wohl das Verständnis des Begriffes „Objektivität“ im Sinne eines Gütekriteriums von Messungen, das von den übrigen Hauptgütekriterien Reliabilität und Validität flankiert bzw. ergänzt wird. Eine Messung gilt als objektiv, wenn mehrere unabhängige Beobachter bei der Anwendung derselben Messvorschrift zum selben Ergebnis gelangen bzw. das Ergebnis einer Messung unabhängig von der Person ist, die das jeweilige Messinstrument anwendet. Das Gütekriterium Objektivität beschreibt somit die Standardisierung von Durchführung, Auswertung und Interpretation von Messungen und lässt sich dabei entlang dieser Aspekte differenzieren. (Bühner (2011), S. 58; Diekmann (2011), S. 249; Hirschle (2015), S. 68; Kubinger (2005), S. 159f; Moosbrugger/Kelava (2012), S. 8; Pawlik (2006), S. 18f)

Durchführungsobjektivität ist gegeben, wenn die Messergebnisse der Probanden nicht davon abhängig sind, welcher Beobachter die Messung durchgeführt hat. Erreicht werden kann dies durch eine Standardisierung der Messbedingungen. Ziel sollte es dabei sein, dass sämtliche Umweltbedingungen, die als Störvariablen fungieren könnten konstant gehalten oder kontrolliert werden sollten. Lediglich aus den unterschiedlichen Probanden sollten sich noch Variationen ergeben. (Bühner (2011), S. 59; Diekmann (2011), S. 249; Hirschle (2015), S. 68; Kubinger (2005), S. 160; Moosbrugger/Kelava (2012), S. 9)

Auswertungsobjektivität liegt dann vor, wenn das Messergebnis der Probanden unabhängig vom Auswerter der Messung ist bzw. wenn die Punktzahl eines Probanden je nach Auswerter nicht in der Vergabe unterschiedlicher Messergebnisse resultiert. Eine Verletzung der Auswertungsobjektivität kann sich durch Verzerrungen ergeben, die infolge von Kodierfehlern auftreten. Insbesondere offene Frageformate können hiervon betroffen sein, da hier oft erst im Nachhinein eine Systematisierung der Antworten entlang eines Kategoriensystems erfolgt, welches u.U. Spielraum für Interpretationen zulässt. Begegnet werden kann mangelnder Auswertungsobjektivität durch die Verwendung detaillierter Auswertungsregeln und deren empirischer Überprüfung. Hierzu prüft man beispielsweise inwiefern die Kodierungen verschiedener Auswerter miteinander übereinstimmen. (Bühner (2011), S. 59f; Diekmann (2011), S. 249; Hirschle (2015), S. 69; Kubinger (2005), S. 160; Moosbrugger/Kelava (2012), S. 9f)

Zuletzt meint *Interpretationsobjektivität*, dass im Rahmen der Interpretation von Messwerten keine individuellen Deutungen stattfinden dürfen bzw. gleiche Messergebnisse von unterschiedlichen Forschern nicht verschiedenartig ausgelegt werden dürfen. Interpretationsobjektivität ist somit gegeben, wenn alle beteiligten Forscher die gleichen Schlussfolgerungen aus den Resultaten der Messung ziehen. Gewährleistet wird dies durch die Festlegung und ausführliche Dokumentation von Interpretationsregeln. (Bühner (2011), S. 60; Diekmann (2011), S. 249; Hirschle (2015), S. 69; Moosbrugger/Kelava (2012), S. 10)

Objektivität im Sinne von Fairness stellt eines der Nebengütekriterien empirischer Untersuchungen dar und beschreibt die Forderung nach Objektivität gegenüber dem soziokulturellen Hintergrund der Studienteilnehmer. So darf die Art und Weise der Messung keine systematische Diskriminierung der Probanden aufgrund ihrer sozialen oder kulturellen Herkunft zur Folge haben. (Bühner (2011), S. 73f; Kubinger (2005), S. 160ff) Daneben spielen

bezüglich dieser Form von Objektivität auch die Erfahrungsunabhängigkeit und Verrechnungsfairness eine Rolle. Erstere betrifft die Benachteiligung von Probanden mit wenig Erfahrung in der Bearbeitung psychologischer Tests gegenüber Probanden mit viel Testerfahrung. (Kubinger (1996), S. 513f) Verrechnungsfairness ist gegeben, „wenn die Testwerte eines Tests die zugrundeliegenden empirischen (Leistungs-)Relationen adäquat abbilden“ (Kubinger (1996), S. 514).

Objektive Merkmalsindikatoren zeichnen sich dadurch aus, dass sie „gesellschaftliche Sachverhalte ungefiltert durch individuelle Wahrnehmungsprozesse und unabhängig von persönlichen Wertungen abbilden“ (Noll (2000), S. 4). Demgegenüber sind die Messergebnisse subjektiver Indikatoren an die persönliche Perspektive, Evaluation oder Präferenz des jeweiligen Befragten gebunden. Beispiele für *objektive Merkmalsindikatoren* können dabei neben sozioökonomischen Aspekten, wie Lebenserwartung, Arbeitslosenquote, Bildungschancen oder Einkommen, auch physiologische Messungen, Reaktionszeiten oder die Anzahl eines bestimmten Typs von Antworten sein. Subjektive Indikatoren beinhalten dagegen im Allgemeinen die Erfassung individueller Einstellungen wie beispielweise Zufriedenheit, Glück, Sicherheitsempfinden, aber auch Präferenzen für politische Parteien oder die Einschätzung der wirtschaftlichen Lage, d.h. prinzipiell Variablen über die ausschließlich der spezifische Befragte Auskunft geben kann. Entsprechend können subjektive Indikatoren ausschließlich durch Befragungen erfasst werden. Trotz ihrer augenscheinlichen Differenz weisen sowohl objektive wie auch subjektive Merkmalsindikatoren gleichermaßen hohe Relevanz und Notwendigkeit für die empirische sozialwissenschaftliche Forschung auf. So werden mitunter objektiv identische Lebenssituationen und -bedingungen von verschiedenen Personen höchst unterschiedlich wahrgenommen, wobei oft die subjektiven Situationsdeutungen eine größere handlungsleitende Nachhaltigkeit aufweisen als dies objektive Situationswahrnehmungen vermögen.⁴ Auf der anderen Seite bestehen insbesondere aufgrund der individuellen bzw. subjektiven Spezifität der Messergebnisse subjektiver Merkmalsindikatoren Zweifel an deren Reliabilität und Validität. Vor dem Hintergrund der Relevanz subjektiver Situationswahrnehmung für das individuelle Handeln sollte aber eher von einer unvermeidlichen und unentbehrlichen symbiotischen Ergänzung der beiden Indikatorenformen ausgegangen werden. (Noll (2000), S. 4ff; Weber/Rammsayer (2012), S. 38)

Objektivität im Sinne von Unverfälschbarkeit stellt ebenso wie *Fairness* ein Nebengütekriterium empirischer Untersuchungen dar und meint die Unmöglichkeit für Probanden, das Ergebnis einer Messung willentlich oder unwillentlich in eine präferierte Richtung zu manipulieren. (Bühner (2011), S. 73; Kubinger (2005), S. 162; Hofmann/Kubinger (2001), S. 298) Es werden dabei prinzipiell zwei Arten von Verfälschung unterschieden: (1) Simulation („fake good“) bzw. das Vortäuschen von Verhalten, das man sonst nicht zeigt, und (2) Dissimulation („fake bad“) bzw. das Verschleiern oder Verbergen von Verhalten, das man normalerweise zeigt. (Bühner (2011), S. 125) Beide Möglichkeiten zur Verfälschung lassen sich in den Bereich der sogenannten reaktiven Effekte einordnen und stellen gleichermaßen eine Bedrohung für die Reliabilität und die Validität empirischer Untersuchungen bzw. deren Resultate dar.

Objektive Tests im Sinne von CATTELL versuchen nun insbesondere die hier letztgenannte Form der Objektivität zu erfüllen, wobei sie auch die übrigen Auffassungen von Objektivität

⁴ Vgl. *Thomas-Theorem*: „If men define situations as real, they are real in their consequences.“ (Endruweit/Trommsdorff (2002), S. 252f, S. 557, S. 703; Esser (1999), S. 59ff)

beinhalten bzw. voraussetzen. Es sei bereits an dieser Stelle erwähnt, dass die nachfolgend beschriebenen Objektiven Tests streng von subjektiven Verfahren, wie Selbstbeurteilungsfragebögen, zu unterscheiden sind. Die Bezeichnung dieser Verfahren als objektive Tests geht vielmehr auf die Erfüllung von Objektivität im Sinne des weiter oben beschriebenen Gütekriteriums zurück und ergibt sich aus dem Format der verwendeten Messvorschriften. (Bornstein/Rossner/Hill/Stepanian (1994); Hundleby (1973), S. 65; Loevinger (1957); Meyer/Kurtz (2006), S. 223; Schmukle/Egloff (2011), S. 74) So geben geschlossene Fragen verschiedene standardisierte und in ihrer Wertigkeit vordefinierte Antwortalternativen vor. Die Probanden sind dazu angehalten, anhand dieser vorgegebenen Antwortoptionen eine Selbstbeurteilung hinsichtlich der interessierenden Zieldimension durchzuführen. Hieraus ergibt sich:

„What is *objective* about such a procedure is that the psychologist administering the test does not need to rely on judgment to classify or interpret the test-taker's response; the intended response is clearly indicated and scored according to a pre-existing key. As a result, however, the necessity for judgment is passed on to the test taker. She must interpret the question, consider her personal characteristics, evaluate herself relative to others as best she can, decide the extent to which the characteristic fits her personality, and then choose whether to honestly convey this information in her response.“ (Meyer/Kurtz (2006), S. 223)

Ein Selbstbeurteilungsfragebogen ist somit insofern als objektiver Test zu bezeichnen, als dass jegliche mit der Fragebearbeitung verbundene Interpretationsleistung vollständig dem jeweiligen Probanden obliegt und seitens des Forschers keine klassifizierende Deutung der Antworten mehr notwendig ist.

2.2.2.5.2 Herleitung und Definition

Ausgangspunkt der Entwicklung Objektiver Tests waren die mit Selbstbeurteilungen verbundenen Limitationen und Schwächen. Zwar stellen Selbsteinschätzungsmaße derzeit den Goldstandard bzw. Königsweg innerhalb des Instrumentariums der empirischen Sozialforschung dar, gleichzeitig sind sie aber auch in besonderem Maße von reaktiven Effekten betroffen, was ihre Messgenauigkeit und damit ihre Validität in Frage stellt. (Ortner/Proyer (2015), S. 144; Proyer/Häusler (2007), S. 539; Santacreu/Rubio/Hernández (2006), S. 56f)

„There are, however, three aspects in which personality questionnaires, and in general, assessment based on self-report by individuals is clearly inadequate, which have, to extent, already been indicated: a) involuntary biases due to factors such as the tendency to acquiescence or the inaccuracy of language; b) voluntary distortions of the answers as a result of faking or social desirability; and c) limitations in predicting behavior in different contexts.“ (Santacreu/Rubio/Hernández (2006), S. 56)

Um die Problematik bewusster Verfälschung von Messergebnissen zu entschärfen, entwickelte CATTELL im Rahmen der Differentiellen Psychologie bzw. der Persönlichkeitsforschung sogenannte Objektive Tests, die subjektive Verfälschungen im Wesentlichen dadurch zu vermeiden versuchen, indem das jeweilige Messprinzip für die Probanden möglichst undurchschaubar gestaltet wird und diese keine Selbstbeurteilung leisten müssen. (Fahrenberg (1964), S. 488; Schmidt/Schwenkmezger (1994), S. 229)

CATTELL unterscheidet in diesem Zusammenhang zunächst zwischen L-Daten (*life*), Q-Daten (*questionnaire*) und T-Daten (*test*). L-Daten bezeichnen Daten, die sich aus der Beobachtung des

2. Theorie: Methodik

Verhaltens von Individuen in deren natürlicher Umgebung ergeben. Diese können dabei in Form von Aufzeichnungen hinsichtlich tatsächlicher Ereignisse auftreten oder die Folge von Fremdbeurteilungen sein. Letzteres bedeutet, dass Personen, die mit dem jeweiligen Individuum gut vertraut sind, eine Beurteilung dessen Verhaltens vornehmen. Problematisch bei solchem Datenmaterial ist allerdings, dass es oft der Objektivität im Sinne des Hauptgütekriteriums nicht genügt. So ist eine systematische Beobachtung natürlicher Lebenssituationen oft nur schwer zu realisieren und auch hinsichtlich der Gewährleistung der Vergleichbarkeit von Beobachtungen verschiedener Individuen können Schwierigkeiten auftreten. (Cattell (1958), S. 286; Cattell/Warburton (1967), S. 7; Herzberg/Roth (2014), S. 30, S. 34f; Santacreu/Rubio/Hernández (2006), S. 55f; Schmidt (1975), S. 17)

Q-Daten ergeben sich in der Regel auf Basis von Fragebögen und stellen eine Selbstbeurteilung des jeweiligen Individuums dar. Die Standardisierung des Erhebungsmodus garantiert dabei einen hohen Grad an Objektivität hinsichtlich Durchführung, Auswertung und Interpretation, was die Vergleichbarkeit der Resultate verschiedener Individuen sicherstellt. Die aus der Selbstbeurteilung resultierende Subjektivität der Angaben der Individuen kann dagegen nachteilige Auswirkungen auf die Reliabilität und Validität von Q-Daten haben. So sind Selbsteinschätzungen oft in hohem Maße anfällig für sowohl bewusste als auch unbewusste Verzerrungen oder sind limitiert hinsichtlich der Vorhersage von Verhalten. Bewusste Verzerrungen können sich dabei infolge von fingierten Antworten oder sozial-erwünschtem Antwortverhalten ergeben, während unbewusste Verzerrungen einer Akquieszenztendenz oder sprachlichen Ungenauigkeiten entspringen können. Darüber hinaus kann ein Mangel an Übereinstimmung zwischen berichtetem und tatsächlichem Verhalten bestehen. In diesem Zusammenhang findet oft auch eine Selbsttäuschung der Probanden statt, die aus einem Streben nach Kohärenz im Antwortverhalten erwächst. So sollen sich die eigenen Aussagen sowohl mit den Beschreibungen anderer Personen als auch dem eigenen Verhalten decken. Werden Inkohärenzen entdeckt, beispielsweise durch Dritte, so erfolgt in der Regel eine Rechtfertigung oder Reformulierung der eigenen Aussage, um wieder Übereinstimmung herzustellen. SANTACREU/RUBIO/HERNÁNDEZ bemerken hierzu, dass auch die Stabilität von Aussagen über die Zeit lediglich eine Bestätigung dieses innerhalb der Selbstauskunft verwurzelten Strebens nach Kohärenz darstellt. CATTELL unterscheidet hinsichtlich der beschriebenen Problematik in sogenannte Q- und Q'-Daten. Dabei bezeichnen Q-Daten Aussagen, die durch Beobachtungen des entsprechenden Verhaltens oder Berichte Dritter über das tatsächliche Verhalten bestätigt werden können. Q'-Daten stellen dagegen Beschreibungen dar, die sich ausschließlich aus der Introspektion des Individuums ergeben und dementsprechend allein durch dieses verifiziert werden können. (Cattell (1958), S. 286; Cattell/Warburton (1967), S. 7; Herzberg/Roth (2014), S. 30ff; Santacreu/Rubio/Hernández (2006), S. 55f; Schmidt (1975), S. 17)

T-Daten ergeben sich auf Basis von Messinstrumenten verschiedener Verfahrensklassen, die im Wesentlichen durch eine Verschleierung des Messprinzips gekennzeichnet sind. Im Gegensatz zu sogenannten direkten Maßen, die eine bewusste Aufmerksamkeit der Probanden bezüglich des zu messenden Konstrukts voraussetzen und dieses unmittelbar und explizit abfragen, ist bei solchen indirekten Maßen keine bewusste Aufmerksamkeit seitens der Probanden erforderlich und die Erfassung des interessierenden Konstruktes geschieht, ohne dass eine ausdrückliche Äußerung bezüglich dieses Konstruktes seitens der Probanden notwendig ist. Zu den Verfahrensklassen, aus denen sich T-Daten ergeben, zählen neben Objektiven Tests auch die bereits beschriebenen projektiven und impliziten Messverfahren.

2. Theorie: Methodik

(Herzberg/Roth (2014), S. 30, S. 33f; Jonas/Stroebe/Hewstone (2007), S. 207; Ortner/Horn/Kersting/Krumm/Kubinger/Proyer/Schmidt-Atzert/Schuhfried/Schütz/Wagner-Menghin/Westhoff (2007), S. 61)

Objektive Tests unterscheiden sich von den übrigen indirekten Maßen in erster Linie dadurch, dass sie Situationen des realen Lebens in standardisierter Form simulieren und anhand der Performanz der Probanden innerhalb dieser Miniatursituationen Rückschlüsse auf deren Einstellungen und Charakteristika ziehen. Dabei sollen sie möglichst wenig Kontrolle darüber haben, welche Informationen und auf welche Art und Weise sie diese bezüglich der interessierenden Zieldimension offenbaren. Die Probanden vollziehen hierbei keine Selbstbeurteilung und auch die Art und Weise, in der ihr Verhalten im Zuge der Messauswertung interpretiert wird, bleibt für sie unklar. Mit anderen Worten: Das einem Objektiven Test zugrundeliegende Messprinzip wird den Probanden nicht offengelegt. Ferner wird Verfälschungssicherheit zusätzlich dadurch gewährleistet, dass die jeweiligen Aufgaben keinerlei Augenscheinvalidität hinsichtlich der Messintention aufweisen. Aufgrund der Tatsache, dass den Probanden nicht unmittelbar bzw. per Augenschein klar ist, welches die Zieldimension des Tests ist, wird so eine zielgerichtete Verfälschung verhindert. (Cattell (1958), S. 286ff; Cattell/Warburton (1967), S. 7; Herzberg/Roth (2014), S. 30, S. 33f; Kubinger (1997), S. 756; Kubinger (2006), S. 39ff; Ortner/Horn/Kersting/Krumm/Kubinger/Proyer/Schmidt-Atzert/Schuhfried/Schütz/Wagner-Menghin/Westhoff (2007), S. 62; Ortner/Proyer (2015), S. 135; Ortner/Schmitt (2014), S. 163; Proyer/Häusler (2007), S. 539; Santacreu/Rubio/Hernández (2006), S. 55f; Schmidt (1975), S. 19)

„An objective test is a procedure for obtaining an individual difference score, based on the response to a specific set of stimuli or sequences of stimuli, such that either the correct implication of the response in question is unknown to the subject or the nature of the response is such that the subject cannot readily modify his response in some desired direction.“ (Hundleby (1973), S. 68)

„*Objektive Tests* (...) sind Verfahren, die unmittelbar das Verhalten eines Individuums in einer standardisierten Situation erfassen, ohne dass dieses sich in der Regel selbst beurteilen muss. Die Verfahren sollen für den Pb keine mit der Messintention übereinstimmende Augenschein-Validität haben. Das kann durch die Aufgabenauswahl oder bestimmte Auswertungsmethoden erreicht werden. Um als *Test* zu gelten, müssen auch die Objektiven Verfahren den üblichen Gütekriterien psychologischer Tests genügen.“ (Schmidt (1975), S. 19)

Ausgehend von den hier aufgetragenen Definitionen, stellt das Kerncharakteristikum Objektiver Tests deren potentielle Immunität gegenüber einer willentlich herbeigeführten Abweichung des individuellen Antwortverhaltens vom wahren Wert dar. Diese Immunität ergibt sich dabei infolge einer erfolgreichen Verschleierung der Messintention und des Messprinzips.

Objektive Tests zeichnen sich dabei durch ein breites Anwendungsspektrum aus:

„Thus our answer to the question, ‚What do objective tests measure?‘ is that they measure theoretical constructs, traits, abilities, hypothetical variables, and, indeed, all manner of variables that are not directly observable yet which can be used in models to account for the variation in the personality measures falling into the various domains.“ (Hundleby (1973), S. 76)

Gleichzeitig weisen sie aber auch hinsichtlich der Konzeptionierung von Aufgaben, der verwendeten Inhalte und des jeweils angewendeten Messverfahrens eine große Variabilität auf. ORTNER/PROYER unterscheiden in diesem Zusammenhang drei verschiedene Klassen von Objektiven Tests (Ortner/Proyer (2015), S. 136ff; Schmidt (2006), S. 26):

- (1) *Leistungsaufgaben bzw. Leistungsmessungen*: Die Probanden werden dazu angehalten eine Leistungsaufgabe so schnell und/oder so sorgfältig wie möglich zu lösen, ohne darüber Bescheid zu wissen, was gemessen wird und in welcher Art und Weise das Ergebnis ausgewertet wird. Das Design der gestellten Aufgabe findet dabei in der Regel keine Entsprechung in Situationen des täglichen Lebens. Die inhaltliche Komponente der Leistungsmessung spielt bei Objektiven Tests dieser Klasse keine Rolle, vielmehr werden ausschließlich anhand der Performanz der Probanden Rückschlüsse auf deren Charakteristika gezogen.

Beispiele für diese Gruppe Objektiver Tests sind die *Objective Achievement Motivation Task (OAT)*⁵ zur Messung der Leistungsmotivation und die *time pressure task* zur Erfassung der Stressresistenz. Im Rahmen der *time pressure task* sind die Probanden dazu angehalten, anhand eines zuvor festgelegten Codierschemas Buchstaben und Symbole so schnell und präzise wie möglich einander zuzuordnen. Die hierzu zur Verfügung stehende Zeit wird dabei schrittweise reduziert. Gemessen wird, ob sich die Zuordnungsgeschwindigkeit und Präzision der Probanden daraufhin verringert oder erhöht. (Ortner/Proyer (2015), S. 136f)

- (2) *Simulationen von Situationen des realen Lebens*: Aufgabe der Probanden ist es, mehr oder weniger komplexe Aufgaben zu lösen, die in Simulationen des täglichen Lebens eingebettet sind. Objektive Tests dieser Klasse sind nicht als bloße Leistungsaufgaben designt, auch wenn die Performanz der Probanden im Fokus des Interesses liegt. Aus dem Verhalten der Probanden innerhalb der simulierten Situationen werden dann Schlussfolgerungen auf die Charakteristika der Probanden abgeleitet.

Beispiele für diesen Typ sind der *Betting Dice Test* bzw. *Roulette Test* oder der *Crossing the Street Test*, deren Zieldimension jeweils die Ausprägung der individuellen Risikobereitschaft der Probanden ist. Die Aufgaben müssen dabei jeweils unter Zeitdruck erfüllt werden. Beim *Betting Dice Test* und dem *Roulette Test* wird entweder ein Würfelspiel oder ein Roulettespiel simuliert, wobei die Probanden auf einen von vier Spielausgängen wetten können. Zielvorgabe für die Probanden ist es, innerhalb von zehn Versuchen möglichst viele Punkte zu erreichen. Jeder der vier möglichen Spielausgänge weist eine andere Eintrittswahrscheinlichkeit auf, wobei diejenige mit der niedrigsten Wahrscheinlichkeit den höchsten Punktegewinn verspricht und umgekehrt. Probanden, die häufiger einen Spielausgang mit niedriger Eintrittswahrscheinlichkeit aber hohem Gewinnpotential wählen, zeigen damit eine höhere Risikobereitschaft an, als Probanden, die eine konservativere Spielstrategie wählen. (Ortner/Proyer (2015), S. 137; Rubio/Hernández/Zaldívar/Márquez/Santacreu (2010); Santacreu/Rubio/Hernández (2006), S. 59ff)

Der *Crossing the Street Test* beinhaltet eine Simulation einer Situation im Straßenverkehr, bei der ein Fußgänger so schnell wie möglich und so sicher wie möglich angeleitet werden soll eine Straße zu überqueren. Sichere Strategien resultieren in längerem Laufweg, höherem Zeitaufwand und niedrigeren Testwerten und zeigen dementsprechend eine geringere Tendenz zu risikoreichem Verhalten auf. (Ortner/Proyer (2015), S. 137; Rubio/Hernández/Zaldívar/Márquez/Santacreu (2010); Santacreu/Rubio/Hernández (2006), S. 61f)

⁵ Die Verfahrensweise der *Objective Achievement Motivation Task (OAT)* wird im Rahmen des Abschnittes „2.2.2.5.5 Zur Validität Objektiver Tests“ vorgestellt.

- (3) *Fragebogentyp*: Objektive Tests dieser Klasse nehmen die Form von Fragebogenitems an und konfrontieren die Probanden mit evaluativen Entscheidungen. Die verwendeten Items weisen dabei eine geringe Augenscheinvalidität auf und erfassen andere Konstrukte, als anhand des Item-Inhaltes seitens der Probanden angenommen wird.

CATTELL, HUNDLEBY und SCHMIDT weisen darauf hin, dass herkömmliche Fragebögen im Allgemeinen bzw. größtenteils nicht zu den *Objektiven Tests* zählen, da dort in der Regel eine Selbstbewertung verlangt wird und Fragebogen systematisch verfälscht werden können. Q-Daten, die vornehmlich als Verhalten und nicht in Form einer Selbstbeurteilung erfasst wurden, machen hierbei allerdings eine Ausnahme und können als T-Daten interpretiert werden. In diesem Sinne stellen Q- und T-Daten eher ein Kontinuum als eine Dichotomie dar. Eine Umsetzung Objektiver Tests in Form eines Fragebogens kann durch die Verwendung subtiler Items realisiert werden, die keine Augenscheinvalidität bezüglich der Zieldimension aufweisen und dementsprechend keine systematischen Verfälschungen zulassen. Außerdem werden oft Auswertungsverfahren angewendet, die nicht den Erwartungen der Probanden entsprechen. (Cattell (1958), S. 286ff; Hundleby (1973), S. 67ff; Ortner/Proyer (2015), S. 134; Schmidt (1975), S. 19f) Beispiele für eine Umsetzung Objektiver Tests im Format von Fragebogen-Items und entsprechende Auswertungsstrategien lassen sich bei CATTELL/WARBURTON finden. Hierbei ist allerdings zu bemerken, dass fragebogenähnliche Objektive Tests, deren Item-Design darauf ausgelegt ist, die interessierende Zieldimension zu verbergen, vornehmlich in der ersten Generation Objektiver Tests zum Einsatz kamen. Die zweite Generation Objektiver Tests verzichtet dagegen weitestgehend auf das Fragebogenlayout. (Cattell/Warburton (1967), S. 196ff, S. 276f; Ortner/Proyer (2015), S. 133f, S. 138f; Proyer/Häusler (2007), S. 540) Ein Beispiel für einen neuen Ansatz im Rahmen dieser Kategorie Objektiver Tests stellt allerdings der *Objective Heuristic Thinking Test (OHTT)* dar. Dieser untersucht, inwiefern Heuristiken Entscheidungsprozesse beeinflussen. Im Fokus stehen dabei die *anchoring heuristic*, bei der Entscheidungen anhand eines anpassbaren Startwertes getroffen werden, die *availability heuristic*, bei der anhand der Häufigkeit eines Gegenstandes abgeschätzt wird, wie einfach es ist, Zugang zu ähnlichen Gegenständen zu erhalten, und die *representativeness heuristic*, nach der die Wahrscheinlichkeit eines Ereignisses anhand dessen Ähnlichkeit mit einer übergeordneten Kategorie und inwiefern es besondere Eigenschaften mit dieser Kategorie teilt, abgeschätzt wird. Jede dieser Heuristiken wird mit einer entsprechenden Skala erfasst. Gegenstand der *anchoring scale* sind kurze Aussagen (z.B.: „Imagine you have a bicycle with 7 gears.“) und Fragen, die jeweils in einer lockeren Beziehung zu der vorangegangenen Aussage stehen (z.B. „How many grams does the bicycle’s shifter weigh?“). Erfasst wurde die Abweichung der Angaben der Probanden von der richtigen Antwort. Im Rahmen der *availability scale* sollten die Probanden die Wahrscheinlichkeit bestimmen, mit der eine bestimmte Person oder Gruppe in ein unwahrscheinliches Ereignis (z.B. Flugzeugabsturz) verwickelt wird. Auf der *representative scale* sollten die Probanden bewerten, ob eine Person Mitglied einer bestimmten Gruppe ist. Dabei war den Probanden jeweils die Mitgliedschaftsquote auf der Ebene der Grundgesamtheit für diese und eine zu unterscheidende Gruppe bekannt. (Jasper/Ortner (2014), S. 193f, S. 196; Ortner/Proyer (2015), S. 138f)

Neben dieser Systematisierung lassen sich auch verschiedene Entwicklungsphasen Objektiver Tests unterscheiden. Hintergrund der ersten Generation Objektiver Tests war CATTELLs Annahme, dass eine umfassende Untersuchung der Persönlichkeit nur auf der Basis der oben

beschrieben Daten-Trias von Selbstbeurteilung (Q-Daten), biografischen Daten (L-Daten) und Testdaten (T-Daten) geschehen kann. Objektive Tests nahmen in dieser Generation vornehmlich die Gestalt von Miniaturesituationen an, anhand derer der Verhaltensausdruck bestimmter Persönlichkeitsmerkmale simuliert werden konnte. Dabei wurden neben Verhaltensmustern, physischer Performanz, psychophysiologischen Daten, kreativen Gestaltungsaufgaben und fragebogenähnlichen Tests auch sogenannte projektive Verfahren eingesetzt. Letztere bilden heutzutage eine gesonderte Klasse indirekter Messmethoden. Die zweite Generation Objektiver Tests ist dagegen eng an die Entwicklung moderner Computertechnologie gebunden und bietet neben einem größeren Variationsspielraum, was das Aufgabendesign anging, auch die Möglichkeit eine Vielzahl weiterer Variablen, wie optische Stimuli, Betrachtungsdauer, Reaktionszeit, Reaktionsgeschwindigkeit oder andere Performanzindikatoren, zu modellieren bzw. zu erfassen. (Ortner/Proyer (2015), S. 133f, S. 136; Proyer/Häusler (2007), S. 540)

2.2.2.5.3 Zur Objektivität von Objektiven Tests

Objektive Tests weisen mindestens eine doppelte Objektivität auf, da sie neben der Vermeidung systematischer subjektiver Verfälschungen auch Objektivität in Form intersubjektiver Übereinstimmung, d.h. Objektivität im Sinne des Hauptgütekriteriums, gewährleisten müssen.

„Used to this higher degree of objectivity, objective means not only that the test performance should be similarly scored by two different psychologists, but also that the test stimulus situation, and the whole mode of response, should be such that the testee himself could not fake the response, or distort it to fit his subjective self-concept or his desire.“ (Cattell/Warburton (1967), S. 15)

Betrachtet man die von SANTACREU/RUBIO/HERNÁNDEZ aufgestellte Liste an Charakteristika, die Objektive Tests erfüllen müssen, so wird deutlich, dass Objektive Tests möglicherweise sogar eine dreiwertige Objektivität aufweisen. Im Einzelnen nennen sie dabei die folgenden Vorgaben, die erfüllt sein müssen: (1) Die Messergebnisse dürfen nicht auf Selbstbeurteilungen basieren; (2) die gestellten Aufgaben dürfen keine besonderen Fähigkeiten oder Fertigkeiten seitens der Probanden erfordern; (3) die Messprozeduren dürfen an die Probanden keine Rückmeldung hinsichtlich ihrer Performanz geben; (4) die Messintention muss maskiert werden; (5) die Möglichkeit zur Manipulation oder Kontrolle der Motivation der Probanden muss gewährleistet sein, um zu garantieren, dass diese die gestellten Aufgaben bearbeiten. (Santacreu/Rubio/Hernández (2006), S. 59) Die Punkte (1), (3) und (4) entsprechen dabei Objektivität im Sinne von Unverfälschbarkeit, während (2) als Objektivität im Sinne von Fairness interpretiert werden kann. So dürfen die Probanden nicht schon bereits durch die Art und Weise der Aufgabenstellung diskriminiert werden, sondern allein das Messergebnis soll Unterschiede zwischen den Versuchspersonen aufzeigen.

Die Tatsache, dass innerhalb Objektiver Tests oft objektive Merkmalsindikatoren zur Erfassung der Performanz der Probanden eingesetzt werden, erhöht auf eine vierwertige Objektivität. (Weber/Rammsayer (2012), S. 38)

2.2.2.5.4 Zur Verfälschungssicherheit von Objektiven Tests

Was die Verfälschungssicherheit Objektiver Tests angeht, so zeigen verschiedene Studien, dass es Probanden weit schwerer fällt, die Ergebnisse Objektiver Tests zu verfälschen, als dies ihnen

im Rahmen anderer Erhebungsmethoden wie projektiven Verfahren oder fragebogengestützten Selbstbeurteilungen gelingt. (Ortner/Proyer (2015), S. 141; Ortner/Schmitt (2014), S. 163)

HÄCKER/SCHWENKMEZGER/UTZ untersuchen die Anfälligkeit der von CATTELL entwickelten *Objective-Analytic Test Battery* (OATB) für systematische Verfälschungen anhand zweier experimenteller Studien mit unterschiedlichen Forschungsdesigns. In der ersten Studie bekommen 70 Strafgefangene die OATB zunächst in einer Auslesesituation zur Bearbeitung vorgelegt. Parallel zu dieser Experimentalgruppe bearbeiten 107 Strafgefangene (Kontrollgruppe) die OATB unter neutralen Bedingungen. In der zweiten Studie bearbeiten 58 Studierende (Kontrollgruppe) die OATB unter neutralen Bedingungen und eine Gruppe von 56 Studierenden (Experimentalgruppe) bearbeitet die OATB unter der Bedingung einer SD-Instruktion, deren Vorgabe es ist, sich möglichst vorteilhaft darzustellen. (Häcker/Schwenkmezger/Utz (1979), S. 9f)

Die Ergebnisse für Studie 1 zeigen, dass zwischen Experimental- und Kontrollgruppe lediglich 14 von 75 Testscores signifikante Abweichungen aufweisen. Eine Betrachtung der Richtung der Abweichungen über alle Testvariablen ergibt allerdings, dass die meisten der Testscores innerhalb der Experimentalgruppe deutliche Abweichungen im Sinne sozialerwünschten Antwortverhaltens zeigen. Insbesondere Testaufgaben, die dem traditionellen Fragebogendesign entsprechen, scheinen vermehrt von dieser Tendenz betroffen zu sein, während perzeptiv-motorische bzw. kognitive Testaufgaben weniger stark betroffen sind. (Häcker/Schwenkmezger/Utz (1979), S. 10, S. 14, S. 20f)

Innerhalb von Studie 2 ergeben sich beim Vergleich von Experimental- und Kontrollgruppe bei 22 von 75 Testaufgabe signifikante Unterschiede im Testscore. Dabei können für sechs Variablen die Ergebnisse aus Studie 1 repliziert werden. Auch hier weisen insbesondere Testaufgaben des Fragebogentyps eine erhöhte Anfälligkeit für „*faking-good*“ auf. (Häcker/Schwenkmezger/Utz (1979), S. 14, S. 20f)

Im Vergleich zu den Leistungstests scheinen die Probanden der Experimentalgruppen bei der Bearbeitung des Fragebogentyps keine einheitliche Meinung bezüglich des Konzeptes sozialer Erwünschtheit zu haben. So ergibt sich für die Leistungstests unter Experimentalbedingung grundsätzlich eine Varianzminderung im Antwortverhalten, während für den Fragebogentyp ohne Ausnahme eine entsprechende Varianzerhöhung zu beobachten ist. (Häcker/Schwenkmezger/Utz (1979), S. 15f)

Ausgangspunkt der Studie von ELLIOT/LAWTY-JONES/JACKSON war die Annahme, dass klassische, auf Selbstbeurteilungen basierende Fragebogen leicht zu verfälschen sind, während Probanden im Rahmen der Bearbeitung Objektiver Tests nicht dazu in der Lage sind, ihre Antworten zu manipulieren. Überprüft wurde dies anhand einer experimentellen Anordnung verschiedener Selbstbeurteilungsmaße und Objektiver Tests, mit deren Hilfe sich die Persönlichkeitsdisposition „Impulsivität“ abbilden lässt. (Elliot/Lawty-Jones/Jackson (1996), S. 335ff)

Als Selbstbeurteilungsmaße wurden die *Eysenck Personality Questionnaire Short Scale* (EPQ-RS), der *IVE Questionnaire* sowie ein speziell für die Studie erstellter Impulsivitäts-Fragebogen eingesetzt. Eines der Objektiv-Maße stellte die Zeit dar, die die Probanden zur Bearbeitung der Fragebogen benötigten. Im Rahmen des zweiten Objektiven Tests sollten die Probanden die Linie eines Kreises so langsam wie möglich nachzeichnen. In beiden Fällen wurde davon

ausgegangen, dass impulsive Menschen weniger Zeit benötigen bzw. die jeweilige Aufgabe schneller erfüllen, als weniger impulsive Menschen. (Elliot/Lawty-Jones/Jackson (1996), S. 336f)

Im Rahmen der Untersuchung wurden die Probanden per Zufall auf eine von drei experimentellen Bedingungen aufgeteilt. Die beiden Versuchsgruppen waren dabei in eine fiktive Bewerbungssituation eingebettet, innerhalb derer die Probanden dazu angehalten waren, die Fragebogen in der Art und Weise auszufüllen, wie ein idealer Kandidat für den jeweiligen Job es tun würde. Die innerhalb der jeweiligen Versuchsgruppe ausgeschriebenen Jobs entsprachen dabei Stereotypen besonders risikobereiter/impulsiver (Aktienhändler/innen) bzw. weniger risikobereiter/impulsiver (Bibliothekare/innen) Personen. Die Probanden der Kontrollgruppe sollten dagegen die Fragebogen möglichst ehrlich beantworten. Unabhängig von der jeweiligen Experimentalgruppe sollten die Probanden bei der Bearbeitung der Fragebogen ihr eigenes Tempo wählen. Die Kreisaufgabe erfolgte im Anschluss an die Bearbeitung der Fragebogen. Zum Abschluss sollten die Probanden ungeachtet der jeweiligen Verfälschungsinstruktion den eigens entwickelten Impulsivitätsfragebogen ausfüllen, um den wahren Impulsivitätslevel der Probanden ermitteln zu können. (Elliot/Lawty-Jones/Jackson (1996), S. 336ff)

Die Ergebnisse der Studie zeigten, dass das Antwortverhalten auf den Fragebogen je nach Versuchsbedingung (Aktienhändler/in, Bibliothekar/in, Kontrollgruppe) unterschiedlich ausfiel, während die Ergebnisse der Objektiven Tests unbeeinflusst von der experimentellen Bedingung blieben. So ergaben sich weder bezüglich der Zeitnahme beim Ausfüllen der Fragebogen noch bezüglich der Zeitnahme beim Folgen der Kreislinie ein signifikanter Unterschied zwischen den Experimentalgruppen. Im Falle der Fragebogenbearbeitungszeit ergaben sich sogar identische Durchschnittswerte. Es lässt sich demnach schließen, dass Selbstbeurteilungsmaße in besonderem Maße anfällig für Verfälschungstendenzen sind, während objektive Maße diesen gegenüber resistent zu sein scheinen. (Elliot/Lawty-Jones/Jackson (1996), S. 339ff)

Als Hauptgrund für diese Resistenz machen ELLIOT/LAWTY-JONES/JACKSON die geringe Augenscheinvalidität der Objektiven Tests und die daraus resultierende Intransparenz der Verbindung mit der Zieldimension „Impulsivität“ verantwortlich. Dagegen begünstigte die für die Selbstbeurteilungsmaße gegebene Augenscheinvalidität Verfälschungstendenzen. (Elliot/Lawty-Jones/Jackson (1996), S. 341)

Darüber hinaus konnte gezeigt werden, dass für die Objektiven Test konkurrente, konvergente, diskriminante Validität und Konstruktvalidität gegeben sind. Es kann somit davon ausgegangen werden, dass beide Objektiv-Maße das gleiche Konstrukt abbilden und jeweils eine gültige Messung der Zieldimension „Impulsivität“ liefern. (Elliot/Lawty-Jones/Jackson (1996), S. 341)

ZIEGLER/SCHMIDT-ATZERT/BÜHNER/KRUMM untersuchen in einer experimentellen Anordnung, ob sich zwischen Selbsteinschätzungen in Form klassischer Fragebogen, projektiven bzw. semi-projektiven Verfahren und Objektiven Tests Differenzen hinsichtlich der Anfälligkeit für Verfälschungstendenzen ergeben. Sie bedienen sich hierzu dreier Vertreter der obengenannten Verfahrensklassen, die jeweils die Zieldimension „Leistungsmotivation“ abbilden können. Einschränkend ist hierbei zu bemerken, dass die einzelnen Messverfahren nur eine vergleichsweise geringe konvergente Validität aufwiesen, sodass möglicherweise unterschiedliche Konstrukte abgefragt wurden. Die Autoren weisen in diesem Zusammenhang allerdings daraufhin, dass dieses Ergebnis im Einklang mit vorangegangenen

Untersuchungsergebnissen steht und keinen Widerspruch zur Konstruktvalidität der Verfahren darstellt. (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 291ff; S. 304)

Im Rahmen der Selbsteinschätzung kommt dabei eine Kurzform des *Achievement Motivation Inventory (LMI-K)* zum Einsatz. Dieses stellt einen insgesamt 30 Statements umfassenden Fragebogen dar, dessen Items von den Probanden anhand einer 7-stufigen Ratingskala bewertet werden können. (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 296)

Repräsentant semi-projektiver Verfahren ist das *Multi-Motive-Grid (MMG)*, welches 18 mehrdeutige Bilder enthält, denen die Probanden jeweils eines von drei vorgegebene Statements zuordnen sollten, die drei verschiedene Motivationsbedürfnisse (Leistung, Macht, Zugehörigkeit) repräsentierten. (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 296)

Als Vertreter Objektiver Tests wurde der *Objective Achievement Motivation Test (OLMT)* ausgewählt. Dieser den Leistungsaufgaben zuordenbare Test umfasst drei Subtests, die verschiedene Aspekte des Konstruktes „Leistungsmotivation“ erfassen sollen (Baseline-Motivation, Motivation durch persönliche Ziele, Anspruchsniveau und Motivation infolge von Wettbewerb). Aufgabe der Probanden ist es, sich entlang eines auf einem Bildschirm dargestellten Pfad zu bewegen, der sich abwechselnd aus rot bzw. grün eingefärbten Abschnitten zusammensetzt. Der Pfad verläuft in Schlangenlinien, wobei grüne Abschnitte eine Linkskurve und rote Abschnitte eine Rechtskurve beschreiben. Die Bewegung auf den verschiedenfarbigen Abschnitten ist jeweils an die Betätigung unterschiedlicher Tasten gebunden. Der erste Subtest besteht darin, dass die Probanden binnen 10 Sekunden eine möglichst große Distanz auf dem Pfad zurücklegen sollen, wodurch die grundlegende Motivation der Probanden geprüft werden soll. Im zweiten Subtest sollen die Probanden im Vorhinein festlegen, wie groß die Distanz sein wird, die sie beim nächsten Versuch erreichen werden. Die Differenz zwischen erreichter Distanz und vorgegebener Distanz dient hierbei als Messwert für das Anspruchsniveau der Probanden. Im dritten Subtest treten die Probanden gegen einen Wettbewerber an, der um 10% schneller ist als diese in ihren jeweils letzten drei Versuchen abgeschnitten haben. Im zweiten und dritten Subtest werden Punkte vergeben, wenn das selbstgesetzte Ziel erreicht oder überschritten wird (Motivation durch persönliche Ziele) bzw. wenn der Wettbewerber geschlagen wird (Motivation durch Wettbewerb). Für Versagen werden jeweils Punkte abgezogen. (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 295)

Die Probanden wurden per Zufall auf eine von drei Gruppen aufgeteilt, wobei der *fake-good*- bzw. der *fake-bad*-Gruppe die Instruktion vorgeschaltet war, sich möglichst positiv bzw. negativ darzustellen. Die Kontrollgruppe bekam keine gesonderte Vorgabe hinsichtlich des Antwortverhaltens. Als fiktives Setting des Experimentes wurde ein berufliches Auswahlverfahren vorgegeben. In jeder der Gruppen mussten alle genannten Messverfahren bearbeitet werden, wobei die Präsentationsreihenfolge dem Zufallsprinzip folgte. Um unrealistische Antworten zu vermeiden, wurden die Probanden außerdem darauf hingewiesen, dass die Resultate auf Verfälschungen hin untersucht werden. (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 296f)

Es zeigte sich, dass alle der eingesetzten Messverfahren anfällig für *faking-bad* waren. Hinsichtlich *faking-good* ergaben sich für den OLMT keine signifikanten Unterschiede zwischen *fake-good*- und Kontrollgruppe, während das LMI-K auch für diese Form der Verfälschung anfällig war. Für das MMG ergab sich je nach betrachteter Konstruktdimension eine Anfälligkeit bzw. keine Anfälligkeit für *fake-good*. Außerdem wurde untersucht, wie die Probanden den Erfolg von Verfälschungstendenzen je nach Messverfahren beurteilen. Hierbei ergab sich, dass im Rahmen des LMI-K *faking-good* und *faking-bad* in gleichem Maße als erfolgreich eingeschätzt

wurden. Hinsichtlich des *OLMT* wurde *faking-good* als schwieriger eingeschätzt als *faking-bad* und bezüglich des *MMG* ergab sich die umgekehrte Relation. Innerhalb der *fake-good*-Gruppe wurde die Verfälschungsperformanz auf dem *OLMT* als signifikant schlechter bewertet als auf *LMI-K* und *MMG*, während sich zwischen den beiden letztgenannten kein Unterschied ergab. In der *fake-bad*-Gruppe wurde dagegen die Verfälschungsperformanz auf dem *MMG* signifikant schlechter bewertet als auf dem *OLMT* und dem *LMI-K*. Zwischen den beiden letztgenannten ergaben sich dagegen keine Unterschiede. (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 298ff)

Ausgehend von diesen Resultaten kommen ZIEGLER/SCHMIDT-ATZERT/BÜHNER/KRUMM zu dem Schluss, dass sich die verschiedenen Messverfahren in ihrer Anfälligkeit für Verfälschung unterscheiden. Während insbesondere das Selbstbeurteilungsmaß und das projektive Verfahren für Verfälschungstendenzen empfänglich sind, kann für den Objektiven Test festgestellt werden, dass zumindest *faking-good* unterbunden werden kann. Insgesamt sind allerdings alle getesteten Verfahren mehr oder weniger für Verfälschungstendenzen empfänglich. Im Falle des Objektiven Tests wird die Aufdeckung der Messintention als Hauptgrund für Verzerrungen im Antwortverhalten angesehen. Hingegen scheint die Verschleierung der Auswertungsstrategie eine wirksame Methode zur Eindämmung von Verfälschungstendenzen darzustellen. Im Rahmen der Selbsteinschätzung waren sowohl die Messintention als auch die Auswertungsstrategie ohne weiteres für die Probanden erkennbar. Entsprechend erwies sich der klassische Fragebogen als besonders leicht zu verfälschen. Die Messintention des semi-projektiven Tests konnte von den Probanden weit weniger leicht identifiziert werden, was allerdings Verfälschungstendenzen nicht verhindern konnte. Insgesamt fiel der Effekt aber niedriger aus als beim Selbstbeurteilungsmaß. (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 302ff)

ARENDASY/SOMMER/HERLE/SCHÜTZHOFFER/INWANSCHITZ untersuchen den Einfluss von Verfälschungstendenzen auf die Dimensionalität und die Durchschnittswerte von Objektiven Tests. Im Fokus steht dabei der *Vienna Risk-Taking Test Traffic (WRBTV)* anhand dessen die Persönlichkeitsdisposition „Risikobereitschaft“ erfasst werden soll. Im Rahmen des Tests werden den Probanden verschiedene Videos von Verkehrssituation (Geschwindigkeitswahl, Überholsituationen, Verhalten an Kreuzungen und verschiedene Wetterbedingungen) präsentiert. Die Probanden bekommen die Videos insgesamt zweimal vorgespielt, wobei sie beim zweiten Durchlauf entscheiden sollen, ab wann ihnen das gezeigte Fahrmanöver zu risikobehaftet wird. Die Latenzzeit wird als Maß für das subjektiv akzeptierte Risiko herangezogen. Der Objektive Test kann damit der Klasse *Simulationen von Situationen des realen Lebens* zugeordnet werden. Im Gegensatz zu anderen Objektiven Tests findet hier allerdings keine Verschleierung der Messintention statt. (Arendasy/Sommer/Herle/Schützhofer/Inwanschitz (2011), S. 210ff)

ARENDASY/SOMMER/HERLE/SCHÜTZHOFFER/INWANSCHITZ führten insgesamt zwei Studien mit experimentellem Charakter durch. Studie 1 untersuchte, inwiefern natürlich auftretende Verfälschungstendenzen die Itemparameterschätzung, die Dimensionalität und die Durchschnittswerte des *WRBTV* beeinflussen. Hierzu wurde der *WRBTV* zwei anhand soziodemografischer Charakteristika parallelisierten Gruppen vorgelegt. Eine der Gruppen bearbeitete den Test unter Laborbedingungen, während die andere Gruppe den Test im Rahmen einer verkehrspsychologischen Untersuchung präsentiert bekam. In beiden Fällen war der *WRBTV* in Wahrnehmungsgeschwindigkeits-, Reaktionszeit- und Aufmerksamkeitsmessungen

eingebettet. Die Ergebnisse zeigten, dass unter beiden Versuchsbedingungen die gleiche Persönlichkeitseigenschaft gemessen wurde. Außerdem ergaben sich keine signifikanten Unterschiede, was die Durchschnittswerte der Latenzzeiten angeht. Aus dieser Messäquivalenz schlossen die Autoren, dass sich die Ergebnisse des *WRBTV* über verschiedene Testsituationen hinweg generalisieren lassen. (Arendasy/Sommer/Herle/Schützhofer/Inwanschitz (2011), S. 212ff)

In Studie 2 sollte untersucht werden, ob die Bearbeitung des *WRBTV* unter verschiedenen Verfälschungsinstruktionen die Messäquivalenz verringert und ob eine weniger eindeutige Beschreibung der idealen Ausprägung der Zieldimension Verfälschungstendenzen wirksam behindern kann. Zunächst bearbeiteten hierzu zwei entlang soziodemografischer Variablen parallelisierte Gruppen den Test unter Standardbedingungen, woran sich mit einwöchigem Abstand die Bearbeitung unter einer von zwei Verfälschungsbedingungen anschloss. Beide Versuchsbedingungen beinhalteten eine fiktive Bewerbungssituation, wobei sich die Probanden so verhalten sollten, dass sie einem möglichst idealen Bewerber entsprachen. Aus den Instruktionen der ersten Versuchsbedingung ging dabei hervor, dass der ideale Jobkandidat eine möglichst geringe Risikobereitschaft aufweist, während die Instruktionen der zweiten Versuchsbedingungen nur ambivalente Hinweise auf die gewünschte Risikobereitschaft gab. Die zweite Versuchsgruppe konnte demnach keine eindeutige Verhaltensempfehlung aus den Instruktionen ableiten. Um die Verfälschungsmotivation der Studienteilnehmer zu erhöhen, wurde die Teilnahme entlohnt, wobei der ausgezahlte Betrag durch eine die Jobchancen maximierende Performanz verdoppelt werden konnte. Aus dem Vergleich der beiden Gruppen ging hervor, dass die drei verschiedenen Versuchsbedingungen (Standard, geringe Risikobereitschaft und Ambivalenz) keinen Einfluss auf die Messäquivalenz hatte. Mit anderen Worten: Der *WRBTV* bildete unabhängig von der Versuchsbedingung die gleiche Dimension ab. Außerdem zeigte sich, dass die Verfälschungsbemühungen in der Ambivalenz-Gruppe erfolgreich behindert werden konnten, während es den Probanden, die die eindeutig formulierte Verfälschungsinstruktion befolgen sollten, leichter fiel, ihre Ergebnisse erfolgreich zu verfälschen. (Arendasy/Sommer/Herle/Schützhofer/Inwanschitz (2011), S. 212ff)

ARENDAZY/SOMMER/HERLE/SCHÜTZHOFFER/INWANSCHITZ schließen aus den Ergebnissen ihrer Studien, dass sich auch Verhaltensmessungen verfälschen lassen. Der Erfolg der Verfälschungsbemühungen ist dabei allerdings vom Scoringverfahren, dem Design der zu erfüllenden Leistungsaufgabe und der Eindeutigkeit der idealen Ausprägung der Zieldimension abhängig. Somit scheint insbesondere die erfolgreiche Verschleierung der Messintention einen besonderen Beitrag zur Verfälschungssicherheit Objektiver Tests zu leisten. (Arendasy/Sommer/Herle/Schützhofer/Inwanschitz (2011), S. 216ff)

Die hier vorgestellten Studien zeigen, dass auch Objektive Tests keinesfalls vollkommen verfälschungssicher sind. Im Vergleich zu anderen Messinstrumenten können Verfälschungseffekte allerdings deutlich und wirkungsvoll vermindert werden. So wird insbesondere *faking-good* bzw. sozialerwünschtes Antwortverhalten durch den Leistungscharakter der jeweiligen Testaufgaben reduziert. Es fällt den Probanden somit schwerer sich positiver darzustellen, als es der Wirklichkeit entspricht. Minderperformanz bzw. *faking-bad* können allerdings noch nicht vollständig ausgeschlossen werden.

Den ungleich größeren Beitrag zur Verfälschungssicherheit von Objektiven Tests scheint allerdings der Aspekt der geringen Augenscheinvalidität der Testaufgaben zu leisten. So behindern die Verschleierung der Messintention oder die Ambivalenz des idealen

Testverhaltens Verfälschungstendenzen nachhaltig. Den Probanden wird es auf diese Weise unmöglich gemacht, ihr Antwortverhalten zielgerichtet zu verändern. Die Aufdeckung der Messintention begünstigt dagegen Verzerrungen im Antwortverhalten.

2.2.2.5.5 Zur Validität von Objektiven Tests

Nicht jedes theoretische Konstrukt kann mit jeder Methode gleich gut erfasst werden. Objektive Tests stehen hierbei oft vor dem Problem, dass eine Zieldimension nicht in eine adäquate Messvorschrift übersetzt werden kann, d.h. eine Operationalisierung scheitert. Unabhängig davon können Objektive Tests allerdings als valide und reliable Alternativen zu direkten und anderen indirekten Messverfahren betrachtet werden bzw. zu deren Ergänzung herangezogen werden, um ein ganzheitliches Bild von einem theoretischen Konstrukt zu erhalten. Nachfolgend sollen nun einige Studien vorgestellt werden, die insbesondere die Validität Objektiver Tests und deren Beziehung zu anderen Messmethoden beleuchten.

COOPER/KLINE/MAY führen eine Kreuzvalidierung einer Auswahl Objektiver Tests aus der *Objective-Analytic Test Battery* (OATB)⁶ durch, die zur Messung von Autoritarismus, Neurotizismus, Psychotizismus und Machiavellismus verwendet werden können. Untersucht wird dabei, inwiefern die ausgesuchten Testaufgaben mit den Ergebnissen verschiedener Fragebogen korrelieren. Zur Erfassung der verschiedenen Konstrukte wurden dabei die nachfolgenden Instrumente verwendet: Wilson-Patterson Attitude Inventory und F-Scale (Autoritarismus), N-Scale des *Eysenck-Personality Questionnaire* (EPQ) (Neurotizismus), P-Scale des EPQ (Psychotizismus) und die Machiavellismus-Skala. Es nahmen insgesamt 42 Studierende an der Studie teil. (Cooper/Kline/May (1986), S. 16ff)

Die Resultate der Untersuchung zeigten, dass die verwendeten Objektiven Tests zur Erfassung der Konstrukte Autoritarismus, Psychotizismus und Machiavellismus durchaus eingesetzt werden können bzw. für diese Zieldimensionen weitestgehend Konstruktvalidität angenommen werden kann. Im Falle von Autoritarismus wiesen drei von fünf der entsprechenden Objektiven Tests ausreichend hohe und signifikante Korrelationen mit den zugehörigen Skalen auf, auch der zur Messung von Psychotizismus eingesetzte Objektive Test korrelierte mittelstark und signifikant mit der P-Scale und vier von acht der zur Erhebung von Machiavellismus benutzten Testaufgaben wiesen adäquate und signifikante Korrelationen mit der zugehörigen Skala auf. Neurotizismus konnte durch die Objektiven Tests dagegen weniger gut abgebildet werden. Zwar ergaben sich signifikante Zusammenhänge mit der N-Scale, allerdings waren die Korrelationen im Betrag eher schwach. (Cooper/Kline/May (1986), S. 18ff)

LEJUEZ/READ/KAHLER/RICHARDS/RAMSEY/STUART/STRONG/BROWN untersuchen inwiefern die *Balloon Analogue Risk Task* (BART) Konvergenzen mit Selbstauskunftsmaßen aufweist, die risikoverwandte Konstrukte erfassen bzw. reales Risikoverhalten abbilden. Es nahmen insgesamt 86 Personen an dieser Studie teil.

Die Selbstauskunftsatterie setzte sich aus einer Vielzahl von unterschiedlich langer Einzelskalen zusammen, die im Wesentlichen Impulsivität, die Affinität zu Spannungsreizen,

⁶ Für eine Konstruktvalidierung der von CATTELL entwickelten *Objective-Analytic Test Battery* (OATB), die eine Selektion von 76 Objektiven Tests aus dem mehr als 800 Objektiven Tests umfassenden Kompendium von Cattell/Warburton (1967) enthält, vgl. Kline/Cooper (1984).

2. Theorie: Methodik

Defizite in der Verhaltenskontrolle sowie Sucht-, Gesundheits- und Sicherheitsverhalten erfassten. (Lejuez/Read/Kahler/Richards/Ramsey/Stuart/Strong/Brown (2002), S. 76f)

Bei der BART bestand die Aufgabe der Probanden darin, einen simulierten Luftballon aufzupumpen, wobei für jeden Pumpvorgang, bei dem der Ballon nicht platzte, ein Geldbetrag auf ein temporäres Konto eingezahlt wurde. Platzte der Ballon, war das gesammelte Geld verloren, wobei das Geld jederzeit auf ein permanentes Konto überwiesen werden konnte. Insgesamt wurden 90 Durchgänge gespielt, wobei ein Durchgang beendet war, wenn entweder der Ballon geplatzt war oder Geld auf das permanente Konto überwiesen worden war. Die Wahrscheinlichkeit, dass ein Ballon platzt, wurde nach jedem Pumpvorgang neu vom Computer berechnet. (Lejuez/Read/Kahler/Richards/Ramsey/Stuart/Strong/Brown (2002), S. 77f)

Es ergaben sich mit wenigen Ausnahme sowohl zwischen der BART und risikoverwandten Konstrukten sowie der BART und selbstberichtetem Risikoverhalten signifikante Korrelationen, deren Betrag allerdings als schwach bis mittelstark beschrieben werden muss. Es kann aber trotz allem von konvergenter Validität ausgegangen werden. (Lejuez/Read/Kahler/Richards/Ramsey/Stuart/Strong/Brown (2002), S. 79ff)

DISLICH/ZINKERNAGEL/ORTNER/SCHMITT untersuchten, inwiefern die Resultate direkter und impliziter Messverfahren sowie Objektiver Tests konvergieren. Als implizite Maße kamen die *Semantic Misattribution Procedure* (SMP), sowie ein klassischer IAT (IAT) und ein *Single-Category IAT* (SC IAT) zum Einsatz. Im Rahmen des SMP bekamen die Probanden zunächst für wenige Millisekunden Worte präsentiert, die entweder mit Risiko oder Sicherheit verbunden waren. Hiernach schloss sich ebenfalls für wenige Millisekunden die Präsentation eines Piktogramms an. Die Probanden sollten nun bestimmen, ob sie das Piktogramm mit einem Risiko oder mit Sicherheit assoziierten. Beim IAT sollten die Probanden die im Rahmen der SMP verwendeten Worte sich selbst oder anderen zuordnen. Beim SC IAT sollten sie die Worte aus der SMP der Kategorie „Glücksspiel“ zuordnen. (Dislich/Zinkernagel/Ortner/Schmitt (2010), S. 22)

Als direkte Messverfahren wurden verschiedene Ratingskalen verwendet. Die *DOSPERT-G* setzte sich aus insgesamt drei Einzelskalen zusammen, welche die Risikowahrnehmung hinsichtlich bestimmter Glücksspiele, die Gewinnerwartung bezüglich dieser Glücksspiele und die Wahrscheinlichkeit an einem dieser Glücksspiele teilzunehmen, abfragten. Im Kontext der *Explicit Rating Scale* sollten die Probanden die im Rahmen der SMP verwendeten Worte auf einer sechsstufigen Ratingskala als risikobezogen oder sicherheitsbezogen bewerten. (Dislich/Zinkernagel/Ortner/Schmitt (2010), S. 22f)

Vertreter der Objektiven Tests waren die *Balloon Analog Risk Task* (BART) und die *Game of Dice Task* (GDT). Die BART wurde analog zu LEJUEZ/READ/KAHLER/RICHARDS/RAMSEY/STUART/STRONG/BROWN umgesetzt und unterschied sich lediglich darin, dass insgesamt nur zehn Durchgänge gespielt wurden. Die Aufgabe der GDT bestand darin, das eigene Startkapital innerhalb eines Würfelspiels mit zehn Durchgängen möglichst zu maximieren. Die Probanden konnten dabei auf eine von mehreren Gewinnoptionen setzen, wobei deren Eintrittswahrscheinlichkeit mit der Steigerung des potentiellen Gewinns abnimmt. (Dislich/Zinkernagel/Ortner/Schmitt (2010), S. 23)

Es ergaben sich keine signifikanten und relativ schwache Korrelationen zwischen den impliziten Maßen. Die Korrelation zwischen den direkten Messverfahren wurde signifikant, erreichte allerdings auch keine hohen Werte. Die verwendeten Objektiven Tests korrelierten schwach negativ miteinander. Bis auf eine signifikant negative Korrelation zwischen BART und

dem SC-IAT ergaben sich zwischen den Objektiven Tests und den impliziten Maßen keine signifikanten Zusammenhänge. Von den Objektiven Tests korrelierte außerdem nur die GDT signifikant mit beiden direkten Messverfahren. Während die BART somit einen Mangel an konvergenter Validität gegenüber Selbstbeurteilungsmaßen aufweist, ergibt sich für die GDT eine geringe Konvergenz mit impliziten Messverfahren. (Dislich/Zinkernagel/Ortner/Schmitt (2010), S. 24f)

RUBIO/HERNÁNDEZ/ZALDÍVAR/MÁRQUEZ/SANTACREU untersuchen die konvergente Validität sowie die Kriteriumsvalidität von Objektiven Tests zur Erfassung von Risikobereitschaft. Zum Einsatz kommen dabei der *Roulette Test* (RT) und der *Crossing the Street test* (CtST). Beim RT besteht die Aufgabe für die Probanden darin, auf einen von vier möglichen Spielausgängen zu wetten. Jeder der Spielausgänge weist eine andere Eintrittswahrscheinlichkeit und eine andere Gewinnmarge auf, wobei sich beide Größen invers zueinander verhalten. Zielvorgabe für die Probanden ist es, möglichst viele Punkte innerhalb von zehn Spieldurchgängen zu erreichen. Beim CtST soll ein Fußgänger so schnell wie möglich und ohne Verursachung eines Unfalls über eine Straße geleitet werden. Als externes Kriterium wird die Veranlagung der Probanden zu Schätzungen anhand von Multiple-Choice-Tests gemessen. (Rubio/Hernández/Zaldívar/Márquez/Santacreu (2010), S. 89f)

Insgesamt werden zwei Studien mit zwei verschiedenen Stichproben und unterschiedlichem Forschungsdesign durchgeführt. In der ersten Studie kamen sowohl der RT als auch der CtST zum Einsatz, wobei sich die Stichprobe aus 1325 Universitätsabsolventen zusammensetzte. In der zweiten Studie kam lediglich der RT zum Einsatz und es nahmen 127 Psychologiestudierende teil. (Rubio/Hernández/Zaldívar/Márquez/Santacreu (2010), S. 90f)

Die Resultate der ersten Studie ergaben eine durchschnittliche Konvergenz zwischen den Objektiven Tests, was darauf hindeutet, dass beide jeweils unterschiedliche Facetten von Risikobereitschaft abbilden. In beiden Studien ergaben sich jeweils signifikante allerdings lediglich schwach bis mittelstarke Korrelationen zwischen den Objektiven Tests und dem externen Kriterium. Nichtsdestotrotz scheint Kriteriumsvalidität gegeben zu sein bzw. mit Hilfe der Objektiven Tests ließ sich das konkrete Verhalten der Probanden voraussagen. (Rubio/Hernández/Zaldívar/Márquez/Santacreu (2010), S. 91ff)

ZIEGLER/SCHMUKLE/EGLOFF/BÜHNER untersuchten, inwiefern direkte und indirekte Maße zur Erfassung von Leistungsmotivation ein gemeinsames Konstrukt abbilden und ob sie sich hinsichtlich der Kriteriumsvalidität unterscheiden. Auf Seiten der indirekten Messverfahren kommen ein IAT als implizites Messverfahren, der *Objective Achievement Motivation Test* (OAMT) als Objektiver Test und ein picture-story test (PST) als projektives Verfahren zum Einsatz. Im Rahmen des IAT bekamen die Probanden Stimuli präsentiert, die entweder selbst- oder fremdbezogen waren (*me*: „I“, „self“, „my“, „own“ vs. *other*: „they“, „your“, „others“, „you“) und denen jeweils Attribute der Kategorien *successful* („ambitious“, „curious“, „persistent“, „diligent“, „inventive“, „efficient“, „successful“, „competent“) und *not successful* („idle“, „uninterested“, „sluggish“, „distractible“, „unimaginative“, „inefficient“, „unsuccessful“, „incompetent“) zugeordnet werden sollten. Der OAMT konfrontiert die Probanden mit der Aufgabe, an einem Bildschirm mit Hilfe der Cursortasten entlang einer gewundenen Straße zu navigieren. Für die direkten Messverfahren kamen die *Achievement Motivation Scale-Revised* (AMS-R) und die *Conscientiousness Facet Achievement Striving* (AS) zum Einsatz. Die AMS-R differenziert sich nochmals in die zwei Subskalen HS (hope for success) und FF (fear of failure)

aus, die jeweils fünf Items umfassen, welche anhand einer fünfstufigen Ratingskala bewertet werden. Die AS umfasst acht Items, die ebenfalls anhand einer fünfstufigen Ratingskala bewertet werden können. Zusätzlich wurden der *NEO-PI-R* und der *Intelligence Structure Test 2000-R* mitlaufen gelassen, um verschiedene Persönlichkeitscharakteristika, wie die Big 5 und Intelligenz, der Probanden zu erfassen. (Ziegler/Schmukle/Egloff/Bühner (2010), S. 16f)

An der Untersuchung nahmen insgesamt 150 Psychologiestudierende teil. Als externes Kriterium wurde deren Abscheiden bei einer Probeklausur bzw. deren Abschneiden bei einer realen Statistiklausur gewählt. (Ziegler/Schmukle/Egloff/Bühner (2010), S. 16f)

Die Resultate der Studie zeigen, dass die indirekten Messverfahren kaum bzw. gar nicht korrelieren, während die direkten Messverfahren zwar schwach, aber signifikant miteinander korrelieren. Es zeigt sich außerdem, dass die direkten Messverfahren offensichtlich ein gemeinsames Konstrukt repräsentieren, während dies für die indirekten Messverfahren nicht gilt. Als Begründung hierfür führen die Autoren die unterschiedlichen Messkonzepte der verwendeten indirekten Verfahren an. Kriteriumsvalidität lässt sich ausschließlich für den OAMT, und die Subskalen der AMS-R nachweisen. Der Objektive Test stellt somit das einzige indirekte Messverfahren dar, welches die Klausurnote der Probanden voraussagen konnte. (Ziegler/Schmukle/Egloff/Bühner (2010), S. 18ff)

JASPER/ORTNER entwickeln und validieren einen Objektiven Test des Fragebogentyps zur Erfassung dreier Strategien heuristischer Entscheidungsprozesse: (1) Verankerung (anchoring), (2) Verfügbarkeit (availability) und (3) Repräsentativität (representativeness). Jeder der Heuristiken wird mit einer separaten Skala⁷ abgefragt. (Jasper/Ortner (2014), S. 193f, S. 196)

An Studie 1 nahmen insgesamt 300 Studierende teil, von denen 105 auch am Retest teilnahmen. Zusätzlich zu den Skalen des *Objective Heuristik Thinking Test* (OHTT) wurden Persönlichkeitscharakteristika, die Schulnoten in den Fächer Deutsch und Mathematik, die Neigung zu Sensationslüsternheit (*Arnett Sensation Seeking Scale*), Ängstlichkeit (*State-Trait Anxiety Inventory*) und verbale Intelligenz (*Verbal Intelligence Scale*) abgefragt. (Jasper/Ortner (2014), S. 196f)

Für die drei Skalen des OHTT ergaben sich nach Elimination einzelner Items akzeptable bis exzellente Werte hinsichtlich deren innerer Konsistenz. Auch die Retest-Reliabilitäten der Skalen sind als sehr gut zu bewerten. Mit Hilfe einer Hauptkomponentenanalyse konnte zusätzlich nachgewiesen werden, dass durch die Subskalen des OHTT drei latente Faktoren repräsentiert werden. Hinsichtlich der Konstruktvalidität ergaben sich gemischte Resultate. Insgesamt blieben die Korrelationswerte mit den Tests ähnlicher bzw. anderer Gültigkeitsbereiche auf einem vom Betrag her eher niedrigen Level. Allerdings ergaben sich teilweise auch signifikante Zusammenhänge. (Jasper/Ortner (2014), S. 197f, S. 200f)

An Studie 2 nahmen insgesamt 55 Studierende teil, wobei neben dem OHTT der Gestaltwahrnehmungstest, der *Intelligence Structure Test 2000 R*, der *Adaptive Matrices Test* (AMT) und der *CRT* zur Erfassung kognitiver Reflektion präsentiert wurden. Auch hier ergaben sich gute bis exzellente Werte für die Ausprägung der inneren Konsistenz bzw. die in Studie 1 nachgewiesene Verlässlichkeit der Subskalen des OHTT konnten repliziert werden. Die Werte für die Konstruktvalidität rangierten ähnlich wie bereits in Studie 1 in einem vergleichsweise niedrigen Bereich. (Jasper/Ortner (2014), S. 199ff)

⁷ Die Gestaltung der Skalen des *Objective Heuristik Thinking Test* (OHTT) wurde bereits im Rahmen des Abschnitts „2.2.2.5.2 Herleitung und Definition“ beschrieben.

KOCH/ORTNER/EID/CASPERS/SCHMITT untersuchten die Reliabilität und Validität Objektiver Tests anhand eines Längsschnittdesigns und im Kontrast zu impliziten Messverfahren und Selbstbeurteilungsmaßen. Zieldimensionen der eingesetzten Messmethoden waren die Konstrukte „Gewissenhaftigkeit“ und „allgemeine Intelligenz“, wobei jedes der Konstrukte mit Hilfe zweier Objektiver Tests, einem IAT und einem Selbstauskunftsfragebogen erfasst wurde. Um eine Konfundierung der Messungen zu vermeiden, wurden die Selbstbeurteilungsmaße erst im Anschluss an die Objektiven Tests und die IATs präsentiert. Die Messungen wurden dreimal im Abstand von einer bis zwei Wochen wiederholt, wobei insgesamt 367 Studierende teilnahmen. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 213ff)

Die zur Messung von Gewissenhaftigkeit eingesetzten Objektiven Tests waren der *Conscientiousness Objective Personality Test* (COPT) und der *Conscientiousness Test 62* (C62). Der COPT beinhaltete die Aufgabe jeden Buchstaben „d“ aus einem vorgegeben Text auszustreichen, wobei auch ähnlich aussehende Buchstaben enthalten waren. Die Anzahl von Fehlern wurde als Indikator für die Zieldimension verwendet. Im Kontext des C62 sollten die Probanden aus einem Figurenpaar die jeweils größere auswählen. Dabei wurde die Anzahl der korrekten Entscheidungen als Indikator herangezogen. Im Rahmen des zugehörigen IAT sollten sich die Probanden Attribute zuschreiben, die für Gewissenhaftigkeit stehen bzw. dieser entgegengesetzt sind (z.B. „accurate“, „careful“, „tidy“ vs. „careless“, „faithless“, „chaotic“). Das Selbstbeurteilungsmaß enthielt insgesamt zehn Items, die auf einer fünfstufigen Ratingskala bewertet werden sollten (z.B. „I think thoroughly through things before making a decision.“). Als Verhaltenskriterium für Gewissenhaftigkeit wurde die Pünktlichkeit, mit der die Probanden zur Studie erschienen, erfasst. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 214)

Zur Erfassung allgemeiner Intelligenz kamen die *Objektiven Tests Number Series* (NST) und *Adaptive Matrices Test* (AMT) zum Einsatz. Beim NST bestand die Aufgabe der Probanden darin, eine Zahlenreihe korrekt weiterzuführen, während der AMT logisches Denken anhand figuraler Items maß. Im zugehörigen IAT sollten sich die Probanden Attribute der Kategorien „intelligent“ bzw. „dumm“ zuschreiben (z.B.: „clever“, „bright“, „able“ vs. „silly“, „dumb“, „foolish“). Der Selbstbeurteilungsfragebogen enthielt selbsterstellte Items, die anhand einer fünfstufigen Ratingskala beantwortet werden sollten (z.B. „I do not consider myself as intelligent.“). (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 215)

Bezüglich der Messung von Gewissenhaftigkeit ergaben sich nur geringe Konvergenzen zwischen Objektiven Tests und den übrigen Maßen. Dagegen wiesen für die Messung allgemeiner Intelligenz insbesondere die Objektiven Tests und das Selbstbeurteilungsmaß eine vergleichsweise hohe Übereinstimmung auf. Mit anderen Worten: Die konvergente Validität der Objektiven Tests war für die Messung allgemeiner Intelligenz stärker ausgeprägt als für die Messung von Gewissenhaftigkeit. Gleiches gilt für die Korrelationen der Objektiven Tests untereinander. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 224f)

Die diskriminante Validität der Objektiven Tests rangierte in einem annehmbaren Bereich. Es kann somit angenommen werden, dass die einzelnen Objektiven Tests jeweils trennscharfe Messungen darstellen. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 225f)

Anhand der Ergebnisse der Objektiven Tests zur Erfassung von Gewissenhaftigkeit ließ sich die Ausprägung des externen Kriteriums Pünktlichkeit voraussagen bzw. zwischen den jeweiligen Objektiven Tests und der Pünktlichkeit der Probanden bestand eine signifikante Beziehung. Mit anderen Worten: Die Objektiven Tests weisen sogenannte Kriteriumsvalidität auf. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 225f)

Die Reliabilität der Objektiven Tests wird von den Autoren als ausreichend beschrieben. Dabei ergeben sich allerdings für den C62 und den AMT relativ geringe Reliabilitätswerte. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 225f)

Aus den hier vorgestellten Studien geht hervor, dass Objektive Tests durchaus über das Potenzial verfügen, trennscharfe Messungen theoretischer Konstrukte bereitzustellen. Insbesondere die Kriteriumsvalidität ist hierbei hervorzuheben. So waren Objektive Tests in der Regel dazu in der Lage, reales Verhalten bzw. die Ausprägungen externer Kriterien vorauszusagen. Gleichzeitig weisen Objektive Tests allerdings einen Mangel an konvergenter Validität auf. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 209; Ortner/Proyer (2015), S. 139; Rubio/Hernández/Zaldívar/Márquez/Santacreu (2010), S. 88) Die geringe Konvergenz mit anderen Messverfahren wird dabei vornehmlich auf zwei Ursachen zurückgeführt. So scheinen die verschiedenen Messmethoden jeweils unterschiedliche Aspekte des gleichen Konstruktes zu erfassen. Dies hängt nicht zuletzt damit zusammen, dass die Messprinzipien von beispielsweise Selbstbeurteilungsmaßen und Objektiven Tests, aber auch die Versuchsanlagen der Objektiven Tests untereinander, mitunter stark voneinander abweichen. (Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 210, S. 224f; Ziegler/Schmukle/Egloff/Bühner (2010), S. 19f)

Daneben scheinen Objektive Tests vornehmlich dazu geeignet zu sein, stabile Merkmale und Eigenschaften von Versuchspersonen zu erfassen. Spontan auftretende Merkmale können dagegen weit weniger gut abgebildet werden. (Jasper/Ortner (2014), S. 201; Koch/Ortner/Eid/Caspers/Schmitt (2014), S. 210, S. 224f; Ortner/Proyer (2015), S. 140)

2.2.2.5.6 Probleme von Objektiven Tests

Im Gegensatz zu anderen indirekten Messverfahren (nicht-reaktive Messverfahren, implizite Messverfahren und projektive Verfahren) haben Objektive Tests in der Vergangenheit kaum nennenswerte Beachtung erfahren und konnten sich nicht zuletzt deshalb bisher nicht nachhaltig durchsetzen. Hierfür ist eine Reihe von Gründen verantwortlich. (Ortner /Schmitt (2014), S. 164)

- (1) *Inkonsistente Benennung*: Objektive Tests firmierten in der Vergangenheit unter einer Vielzahl von Bezeichnungen, wie *Cursive Miniature Situations* (Cattell (1941); Cattell (1944)), *Objective Tests* (Cattell (1946)), *Objective-Analytic Tests* (Cattell (1955)), *Objective Behaviour Tests* (Eysenck (1960a); Eysenck (1960b)), *Performance Tests of Personality* (Cronbach (1970)), *Getarnte Messverfahren* (Petermann/Noack (1999)), *Behavioral Measures* (Lejuez /Read/Kahler/Ramsey/Stuart/Strong/Brown (2002)) und *Experimental-Based Assessments* (Kubinger (2009)). Diese Inkonsistenz in der Benennung hatte zur Folge, dass auch die Dokumentation der entsprechenden Verfahren uneinheitlich ist bzw. die gleichen Verfahren unter verschiedene Oberbegriffe eingeordnet wurden. (Hundleby (1973), S. 68; Ortner/Schmitt (2014), S. 164)
- (2) *Einordnung in das Methodeninstrumentarium*: Auch hinsichtlich der Abgrenzung Objektiver Tests von anderen indirekten Messverfahren bzw. deren Verständnis als eigenständige Klasse von Messverfahren existieren Inkonsistenzen. So bilden Objektive Tests und projektive Verfahren bei CATTELL eine Einheit, die unter dem Begriff *Objective Personality Tests* zusammengefasst wird. (Cattell (1958), S. 285) Auch SCHMIDT/SCHWENKMEZGER

- subsumieren projektive Verfahren unter Objektive Tests. (Schmidt/Schwenkmezger (1994), S. 229) PETERMANN/NOACK bezeichnen Objektive Tests als getarnte Messverfahren und unterscheiden sie dabei von anderen indirekten Messverfahren wie physiologischen Messungen, projektiven Verfahren oder nicht-reaktiven Messverfahren. (Petermann/Noack (1999), S. 440f) Bei ORTNER et al. werden implizite Messverfahren zu den Objektiven Tests gezählt. (Ortner/Horn/Kersting/Krumm/Kubinger/Proyer/Schmidt-Atzert/Schuhfried/Schütz/Wagner-Menghin/Westhoff (2007), S. 61) SCHMUKLE/EGLOFF setzen Objektive Tests mit physiologischen Messungen, projektiven Verfahren und impliziten Maßen gleich und plädieren dafür, diese Klasse von Messverfahren als *indirekte Verfahren zur Messung von Persönlichkeit* zu bezeichnen. Der Term Objektive Tests kann hier somit als Beschreibung für die Gesamtheit indirekter Messverfahren interpretiert werden. (Schmukle/Egloff (2011), S. 73ff) ORTNER/SCHMITT trennen Objektive Tests streng von sogenannten indirekten Messansätzen ab, die allerdings den impliziten Messverfahren entsprechen. (Ortner/Schmitt (2014), S. 163) Bei HERZBERG/ROTH bilden Objektive Tests eine unabhängige Klasse indirekter Messverfahren neben projektiven Verfahren und impliziten Maßen. (Herzberg/Roth (2014), S. 33f)
- (3) *Diversität der Konzeptualisierungen*: Objektive Tests weisen hinsichtlich ihres Aufgabendesigns, der verwendeten Messprinzipien und der Bewertungsregeln für Probandenantworten erhebliche Heterogenität auf. So werden im Rahmen von Objektiven Tests bspw. Reaktionszeiten bzw. -geschwindigkeiten erfasst oder die Aufgabe besteht darin, einen Luftballon aufzupumpen bzw. eine Figur durch ein Labyrinth zu führen. In anderen Fällen sollen Aussagen evaluiert werden oder Geld zwischen sich selbst und anderen aufgeteilt werden. (Ortner/Schmitt (2014), S. 164)
- (4) *Aufgaben- und Materialspezifität sowie Konstruktlimitation*: Objektive Tests sind in ihrer Untersuchungsanlage meist auf die Erfassung eines spezifischen Konstruktes beschränkt und sind dementsprechend weniger adaptionsfähig und flexibel als andere indirekte Messverfahren. Außerdem kann nicht jede Zieldimension ohne weiteres mittels Objektiver Tests erfasst werden. Oft ist der Aufwand verglichen mit alternativen Messmethoden zu groß oder die Messung per Objektivem Test ist vergleichsweise schwer zu implementieren. (Ortner/Schmitt (2014), S. 164; Pawlik (2006), S. 19f)
- (5) *Bestimmung von Reliabilität und Validität*: Im Vergleich zu anderen Messverfahren, wie Selbstbeurteilungsfragebogen, sind Objektive Tests durch eine Vielfalt unterschiedlicher Messmethoden und eine Intransparenz von Messintention und Messprinzip gekennzeichnet. Hieraus ergeben sich Schwierigkeiten hinsichtlich der Bestimmung von Reliabilität und Validität. Es wird deshalb noch vermehrt auf Verfahren zurückgegriffen, deren Verlässlichkeit und Gültigkeit weitestgehend und nachhaltig bestätigt sind. (Ortner/Schmitt (2014), S. 164)
- (6) *Geringe Konvergenz mit anderen Datentypen*: Die Ergebnisse Objektiver Tests weisen teilweise nur eine gering ausgeprägte Übereinstimmung mit den Resultaten von beispielweise Selbstbeurteilungsmaßen auf. Dieser Umstand ist allerdings wohl kaum mit einem Mangel an Validität gleichzusetzen, da nicht wenige Objektive Tests durchaus mit Q-Daten konvergieren und prädiktive Validität aufweisen. (Dislich/Zinkernagel/Ortner/Schmitt (2010); Ortner/Schmitt (2014), S. 164; Proyer (2007))
- (7) *Intransparenz der Messintention*: Die Verschleierung der Zieldimension stellt eine der Kerncharakteristika Objektiver Tests dar und bewirkt im Wesentlichen deren geringe

Anfälligkeit für eine willentliche Verfälschung des individuellen Antwortverhaltens seitens der Probanden. Die Intransparenz der Messintention kann allerdings auch einen Validitätsverlust hervorrufen. Unabhängig davon, ob den Probanden das interessierende Konstrukt bekannt ist, stellen diese Überlegungen hinsichtlich der der Messung zugrundeliegenden *demand characteristics* an. Mit anderen Worten: Die Probanden versuchen herauszufinden, auf welche Art und Weise und was mit dem jeweiligen Test gemessen werden soll. Bleibt die Messintention durch das Messverfahren bzw. die Messinstruktionen unbestimmt, so orientieren sich die Antworten der Probanden an dem, was sie aus den zur Verfügung stehenden Informationen als Zieldimension ableiten können. Hinsichtlich dieser Interpretationsleistung können sich zwischen den Probanden mehr oder weniger große Unterschiede ergeben, sodass nicht jeder das gleiche Konstrukt als Zieldimension identifiziert. Die Folge sind Validitätsbedrohungen, die sich aus Antwortdivergenzen ergeben. Im Wesentlichen können dabei die folgenden Konstellationen auftreten: (a) Die Zieldimensionen von Test und Probanden sind nicht identisch, sodass die Antworten der Probanden nicht mit dem eigentlichen Zielkonstrukt korrespondieren, sondern etwas Anderes abgebildet wird. (b) Jeder einzelne Proband hat eine andere Messintention abgeleitet, sodass neben der eigentlichen Zieldimension noch eine Vielzahl weiterer, abweichender Konstrukte abgebildet wird. (c) Es tritt eine Kombination der obengenannten Problematiken auf. (Bungard (1984), S. 117; Hundleby (1973), S. 68; Orne (1962), S. 779; Ortner/Horn/Kersting/Krumm/Kubinger/Proyer/Schmidt-Atzert/Schuhfried/Schütz/Wagner-Menghin/Westhoff (2007), S. 63)

2.3 Fazit und Ausblick

Sowohl die Verwendung direkter Messverfahren als auch die Verwendung indirekter Messverfahren ist gleichermaßen mit Vor- und Nachteilen verbunden.

So zeichnen sich direkte Messmethoden dadurch aus, dass sie einfach und schnell einsetzbar sind und dabei eine hohe Adaptionfähigkeit aufweisen. Sie ermöglichen eine spezifische Nachfrage hinsichtlich der interessierenden Zieldimension und sind nachweislich reliabel und valide. Gleichzeitig macht ihr vornehmliches Design als Selbstbeurteilungs- und Selbstauskunftsmaße eine bewusste Aufmerksamkeit seitens der Probanden für die jeweilige Messintention notwendig. Hieraus können sich eine Reihe von Problemen ergeben.

Voraussetzung für eine adäquate und qualifizierte Antwort ist dabei zunächst ein Verständnis für die *demand characteristics* bzw. den Anforderungscharakter des jeweiligen Messinstrumentes. Die Probanden müssen zu einer korrekten Interpretation der Messintention in der Lage sein, um im Rahmen einer Selbsteinschätzung eine Antwort geben zu können. Zusätzlich sind manche Aspekte eines Merkmales oder eines theoretischen Konstruktes oft gar nicht aktiv zugänglich oder explizit artikulierbar bzw. ein Proband weiß gar nicht, ob er über die entsprechende Eigenschaft verfügt. Mit anderen Worten: Die Selbstkenntnis der Probanden ist nicht ausreichend stark genug ausgeprägt, als dass sie als kompetente Ansprechpartner bezüglich der interessierenden Zieldimension angesehen werden können. Nicht zuletzt macht die Transparenz der Messintention direkte Messverfahren in besonderem Maße anfällig für reaktive Effekte, aus denen sich Messverzerrungen ergeben können, die wiederum in einer Bedrohung oder Einschränkung der Validität resultieren können.

Der Hauptvorteil indirekter Messverfahren liegt in ihrer geringen Anfälligkeit für reaktive Effekte. Diese ergibt sich vornehmlich daraus, dass sie keine bewusste Aufmerksamkeit der Probanden für den Messprozess und die jeweilige Zieldimension voraussetzen. Außerdem spielen hierbei die Verschleierung der Messintention, die weitgehende Vermeidung von Kontakt zwischen Forscher und Probanden und die Datenerhebung innerhalb der natürlichen Lebenswelt der Probanden eine Rolle für die Vereitelung vorsätzlicher Anpassungen des Antwortverhaltens. Indirekte Messverfahren sind dazu in der Lage, Aspekte einer Zieldimension zu erfassen, die mit Hilfe direkter Messverfahren nicht registriert werden können. Ebenso können Phänomene, die nicht verbalisiert werden können, bzw. Einstellungen und Eigenschaften, die den Probanden gar nicht bewusst sind, abgebildet werden.

Nachteile indirekter Messverfahren ergeben sich daraus, dass nicht jedes Phänomen ohne weiteres operationalisiert werden kann bzw. der Aufwand für die Umsetzung einer entsprechenden Messvorschrift zu groß ist. Auch kann sich eine exakte Interpretation des Datenmaterials, aufgrund einer möglichen Ambivalenz der eingesetzten Stimuli schwierig gestalten. Außerdem sind die vielen unterschiedlichen Messkonzepte und -prinzipien oft nur auf spezifische Konstrukte anwendbar und weisen eine relativ geringe Flexibilität auf. Zusätzlich können sich Einschränkungen hinsichtlich Reliabilität und Validität ergeben, die je nach verwendetem Messverfahren unterschiedliche Ursachen haben. So spielt es eine Rolle, ob stabile oder spontan auftretende Eigenschaften und Merkmale gemessen werden. Indirekte Messverfahren erfassen oft andere Facetten eines Konstruktes als Parallelmessungen mit alternativen Messmethoden, sodass sich keine bzw. nur schwache Konvergenzen ergeben. Die Beschaffenheit der Datengrundlage macht Mehrfachmessungen unmöglich bzw. Parallelmessungen sind zu auffällig, sodass überhaupt keine Vergleiche mit anderen Messverfahren angestellt werden können.

Betrachtet man die hier zusammengefassten wichtigsten Vorzüge und Defizite der verschiedenen Messmethoden, so wird klar, dass eine wechselseitige Ergänzung bzw. parallele Nutzung direkter und indirekter Messverfahren im Rahmen der Datenerhebung es ermöglichen würde, ein holistisches Abbild theoretischer Konstrukte zu erhalten, da die verschiedenen Verfahrensklassen jeweils unterschiedliche Facetten der gleichen Zieldimension erfassen. Viel wichtiger ist allerdings, dass indirekte und direkte Messverfahren anscheinend wechselseitig über das Potenzial verfügen, die methodischen Probleme der jeweils anderen Methodenklasse zu kompensieren. Es erscheint daher nur folgerichtig, eine Kombination beider Verfahrenstypen in Form einer Hybridklasse anzustreben. Ein erfolgversprechender Ansatzpunkt stellt hierbei insbesondere der Fragebogentyp der Objektiven Tests dar, was nicht zuletzt auf seine konzeptionelle Nähe zu traditionellen Selbstbeurteilungs- und Selbstauskunftsmaßen zurückzuführen ist. Der Fragebogentyp erlaubt eine spezifische Nachfrage bei gleichzeitiger Intransparenz der Messintention und vereinigt auf diese Weise die wichtigsten Vorzüge indirekter und direkter Messverfahren. So bewirkt eine Verschleierung der Messintention nachweislich eine Reduktion der Anfälligkeit für reaktive Effekte, indem eine willentliche und vorsätzliche Modifikation des Antwortverhaltens verhindert wird, aus der eine Verfälschung der Testergebnisse resultieren könnte. Eine Hybridklasse sollte somit subtile Items verwenden, die nach Möglichkeit keine mit der jeweiligen Messintention übereinstimmende Augenscheinvalidität aufweisen und auf diese Weise eine bewusste Aufmerksamkeit der Probanden für die Zieldimension des Messinstrumentes unnötig machen.

Welche Gestalt eine solche Hybridklasse aus indirekten und direkten Messverfahren annehmen kann, soll Gegenstand der weiteren Ausführungen sein bzw. am Beispiel des latenten

theoretischen Konstruktes „*generalisiertes Vertrauen*“ entwickelt werden. Die Entwicklung eines unterschwellig wirksamen Messinstrumentes bestehend aus schwer zu durchschauenden Messvorschriften macht es allerdings erforderlich, umfassende und detaillierte Kenntnisse über die theoretischen und empirischen Charakteristika eines Forschungsgegenstandes sowie dessen Verbindungen zu anderen Konstrukten zu besitzen. Nur mit diesem Wissen lassen sich Items designen, die einen indirekten Zugang zu der jeweiligen Zieldimension ermöglichen. Bevor nun die Spezifika des Entwurfs für einen solchen Hybriden⁸ vorgestellt werden, sollen deshalb in den nachfolgenden Kapiteln zunächst die Eigenschaften, Merkmale und Korrelate *generalisierten Vertrauens* aufgearbeitet und definiert werden sowie die klassische Messinstrumente zur Erfassung dieses Konstruktes und deren spezifische methodische Probleme diskutiert werden.

⁸ Vgl. Kapitel „5. Hypothesen“

3. Theorie: Vertrauen

Das theoretische Konstrukt Vertrauen lässt sich je nach Bezugssphäre auf einer grundlegenden Ebene in sogenanntes soziales und politisches Vertrauen differenzieren. Soziales Vertrauen beschreibt im Wesentlichen ein „Vertrauen der Menschen zu ihrer sozialen Umwelt“ (Kunz (2004), S. 202), dagegen sind die Bezugspunkte politischen Vertrauens spezifischer und umfassen im Speziellen politische Institutionen und Personen, die Ämter innerhalb dieser politischen Institutionen innehaben bzw. gewählte Repräsentanten des Volkes darstellen. (Bierhoff (2002), S. 242; Gabriel (1999), S. 202; Göhler (2002), S. 224)

Sowohl politisches als auch soziales Vertrauen lassen sich in jeweils zwei Subkategorien aufspalten. So gliedert sich politisches Vertrauen in sogenanntes Akteurs- und Institutionenvertrauen. (Schweer (2000), S. 11f; Zmerli (2004), S. 230f)

„Vertrauen in der Politik bezieht sich also einerseits auf Personen in sozialen Interaktionen oder auf Personen als Akteure von Institutionen, soweit sie selbst als Repräsentanten gewählt wurden – diesen wird persönlich vertraut oder mißtraut. Andererseits gilt das Vertrauen politischen Institutionen, und zwar den Mechanismen von Institutionen dahingehend, daß diese Mechanismen greifen, auch wenn den Akteuren selbst nicht unmittelbar vertraut wird.“ (Göhler (2002), S. 227)

Beide Spielarten des politischen Vertrauens können aufgrund der Tatsache, dass „Institutionen letztlich immer ihre Akteure haben“ (Göhler (2002), S. 225) kaum losgelöst voneinander betrachtet werden. So konnte WALTER-ROGG zeigen, dass zwischen Akteurs- und Institutionenvertrauen sogenannte „*Spillover*“-Effekte existieren. Die verschiedenen Ausformungen politischen Vertrauens, im Detail sind dies Politikervertrauen, Parteienvertrauen, Vertrauen zu Entscheidungsinstitutionen und Vertrauen zu Implementationsinstitutionen, weisen hier Wechselbeziehungen auf, die keiner hierarchischen Ordnung folgen. Vielmehr ergeben sich Ausstrahlungseffekte in verschiedene Richtungen, die jeweils die Ausprägung und die Veränderung im Vertrauen gegenüber den anderen Objekten erklären. (Walter-Rogg (2005), S. 161, S. 175, S. 178f)

Soziales Vertrauen lässt sich in *partikulares* und *generalisiertes Vertrauen* ausdifferenzieren (Newton/Zmerli (2011), S. 170ff), wobei das Verhältnis zwischen diesen beiden Formen sozialen Vertrauens und insbesondere die Rolle *generalisierten Vertrauens* innerhalb eines Gesellschaftsgefüges Gegenstand der nachfolgenden Erörterungen sein soll.

„Following the standard Oxford English Dictionary definition, the word ‚particular‘ is used here as an adjective describing the noun ‚trust‘ where social trust is associated with specific people or groups of people, whether known or in-group others. General trust is not limited in this way. It extends in a more abstract manner to people as a whole in an unselective and unspecific manner. General trust is inclusive. General trust is the belief that most people can be trusted, even if you do not know them personally, and even if they are not like you socially.“ (Newton/Zmerli (2011), S. 171)

Hinsichtlich der Beziehung von sozialem und politischem Vertrauen existieren innerhalb der Literatur widerstreitende Ansichten. Ausgehend von GABRIEL gilt politisches Vertrauen „als ein Spezialfall sozialen Vertrauens und bezeichnet wie dieses die Disposition von Akteuren, anderen Akteuren bestimmte Handlungsressourcen (...) zur Verfügung zu stellen“ (Gabriel (2002), S. 494). Nach NEWTON/ZMERLI lassen sich aus dem sozialwissenschaftlichen Forschungsstand dagegen bis zu drei verschiedene Modelle für den Zusammenhang von sozialem und politischem Vertrauen ableiten. Das *compatible model* geht davon aus, dass sich die verschiedenen Formen

des Vertrauens gegenseitig verstärken und wechselseitig austauschbar sind. Die unterschiedlichen Vertrauentypen stellen demnach lediglich Varianten ein und derselben Kategorie dar, sodass jemand, der auf einer der Vertrauensebenen vertraut, in der Regel auch auf allen anderen Ebenen vertraut. Im Rahmen der Vertrauensgenese werden hier mehrere Mechanismen wirksam. Auf der einen Seite ergibt sich Vertrauen aus sozialen und politischen Strukturen bzw. Institutionen. Enge soziale Netzwerke und Institutionen, die vertrauenswürdiges Verhalten fördern, generieren gleichermaßen soziales wie politisches Vertrauen, welches positiv auf das gesellschaftliche System zurückwirkt, sodass sich eine reziproke Beziehung ergibt. Auf der anderen Seite stellt Vertrauen ein Kerncharakteristikum der Persönlichkeit von Individuen dar, das sich aus der jeweiligen Primärsozialisation ergibt. Als solches ist es eng verbunden mit weiteren Persönlichkeitsaspekten, wie einem Gefühl für die eigene Selbstwirksamkeit, dem Glaube an interpersonelle Kooperation und einem optimistisches Naturell. (Newton/Zmerli (2011), S. 173f)

Das *incompatible model* geht davon aus, dass die verschiedenen Vertrauentypen voneinander unabhängig sind und nicht ineinander überführt werden können. Hiernach wird soziales Vertrauen durch soziale Aspekte, wie soziale Herkunft, Bildung und Mitgliedschaft in sozialen Netzwerken, beeinflusst, während politisches Vertrauen mit politischen Faktoren, wie politischem Interesse, Parteiidentifikation und Nutzung politischer Medien, assoziiert ist. (Newton/Zmerli (2011), S. 174f)

Das *conditional model* geht davon aus, dass die verschiedenen Varianten des Vertrauens im Prinzip keine Inkompatibilität aufweisen, die Ausrichtung des jeweiligen Verhältnisses aber in hohem Maße von den äußeren Umständen abhängig ist. Hierbei ist allerdings zu bemerken, dass im Rahmen dieses Modells insbesondere der Zusammenhang von *partikularem* und *generalisiertem Vertrauen* im Fokus steht. So wird davon ausgegangen, dass in *generalisiertem Vertrauen* notwendigerweise *partikulares Vertrauen* enthalten sein muss, die umgekehrte Beziehung aber nicht zwingend Gültigkeit besitzt. Gleichzeitig wird angenommen, dass vollkommenes Misstrauen ebenso selten auftritt wie vollkommenes Vertrauen. So zeigt sich, dass eine ausgeprägte Identifikation mit der eigenen *ingroup* keine Feindseligkeit gegenüber der jeweiligen *outgroup* determiniert, sondern eine Funktion des Wettbewerbs um Ressourcen, der gefühlten Bedrohung der *ingroup* durch die *outgroup* und den Vorteilen einer Öffnung gegenüber den Nutzen einer Schließung der *ingroup* hinsichtlich anderer Gruppen darstellt. Außerdem können Veränderungen im kulturellen und institutionellen Kontext, beispielsweise ausgehend von Kriegen und Krisen, eine Verschiebung der Grenzen zwischen *in-* und *outgroup* bewirken, die wiederum einen Effekt auf die Verteilung des Vertrauens haben. Prinzipiell ist es innerhalb liberaler und egalitärer Gesellschaften leichter, umfassendes Vertrauen, d.h. sowohl soziales als auch politisches Vertrauen, zu entwickeln, als dies in durch totalitäre Regime geprägten Gesellschaften der Fall ist. (Newton/Zmerli (2011), S. 175f)

Im weiteren Verlauf der theoretischen Erörterung des Vertrauensbegriffs soll nun das soziale Vertrauen und im Speziellen dessen Subkategorie *generalisiertes Vertrauen* im Fokus stehen.

3.1 Soziales Vertrauen

„Vertrauen kann (1) auf verschiedenen Ebenen auftreten, gegenüber Personen, Organisationen oder Systemen, (2) auf unterschiedlichen Voraussetzungen beruhen, auf Kosten-Nutzenüberlegungen, auf wahrgenommenen Personeneigenschaften oder entstandenen Gemeinsamkeiten, (3) bewusst oder unbewusst mit dem Eingehen eines

3. Theorie: Vertrauen

Risikos verbunden sein, (4) spezifisch oder generell sein und (5) sich auf die Absichten oder die Kompetenz einer Person beziehen.“ (Oswald (2006), S. 710)

Wie sich zeigt, stellt *Vertrauen* keineswegs ein einheitliches Konzept dar. Vielmehr existieren verschiedene Definitionen mit jeweils unterschiedlichen Schwerpunktsetzungen. Entsprechend dieser Heterogenität des Vertrauensbegriffes, gestalten sich auch die existierenden Systematisierungen der verschiedenen Ansätze zur Bestimmung und Beschreibung von sozialem Vertrauen höchst unterschiedlich. Es stellt sich daher zunächst die Frage, welche eine ökonomische und funktionale Strukturierung der unterschiedlichen Vertrauensansätze darstellt, die dabei gleichzeitig der Komplexität des Vertrauensbegriffes gerecht wird. Um diese Frage zu beantworten, sollen zunächst einige exemplarische Beispiele für mögliche Systematisierungen vorgestellt werden, aus denen dann die dieser Arbeit zugrundeliegende Gliederung des Vertrauensbegriffs entwickelt werden wird.

JONES unterscheidet in *trust-as-reliance accounts*, *risk-assessment accounts*, *cognitive accounts* und *goodwill-based accounts*.

Innerhalb der *trust-as-reliance accounts* wird Vertrauen damit gleichgesetzt, sich auf etwas zu verlassen. Objekt solchen Vertrauens kann dabei jedes Lebewesen und jeder Gegenstand sein, der in die Relation „A trusts B to do Z just in case A relies on B’s doing Z“ (Jones (2013), S. 5210) eingefügt werden kann. Vertrauen solcher Art beinhaltet in der Regel keine Risikokomponente. Als Vertrauensobjekte werden nach Möglichkeit nur solche Lebewesen oder Gegenstände ausgewählt, auf die man sich sicher verlassen kann, wobei auch ständig nach weniger riskantem Ersatz für bereits bestehende Vertrauensbeziehungen gesucht wird. (Jones (2013), S. 5210)

Risk-assessment accounts verstehen Vertrauen als eine Wahrscheinlichkeitsabschätzung dessen, ob der Vertrauenspartner ein gewünschtes Verhalten zeigen wird oder nicht. Es werden dabei nur solche Vertrauensbeziehungen eingegangen, bei denen das Risiko eines Vertrauensbruches möglichst gering ist. Diese Risikoabschätzung geschieht dabei entlang von Annahmen über Motive, weshalb der Vertrauenspartner in die Vertrauensbeziehung einsteigt. Es gilt dabei jeder als vertrauenswürdig, der auf dieser Basis als in der Lage angesehen wird, das gewünschte Verhalten zu zeigen. (Jones (2013), S. 5210f)

Cognitive Accounts gehen davon aus, „that the other act from reasons that make reference to the truster“ (Jones (2013), S. 5211). Grundlage solchen Vertrauens ist somit die Annahme, dass beim Vertrauenspartner ein Interesse besteht, dass die Vertrauensbeziehung Bestand hat und er dementsprechend die Erwartungen des Vertrauenden erfüllen möchte. (Jones (2013), S. 5211)

Goodwill-based accounts unterstellen dem Vertrauten ein Wohlwollen gegenüber dem Vertrauenden, das sein Handeln lenkt. Die Handlungen und Motive des Vertrauten werden dabei vom Vertrauenden aus der Perspektive des erwiesenen Vertrauens interpretiert und werden dadurch anders wahrgenommen als dieselben Handlungen einer Person, der nicht vertraut wird. (Jones (2013), S. 5211f)

HARDIN differenziert zwischen *trust as encapsuled interest*, *trust as dependent on characteristics of the trusted*, *trust as a good*, *developmental accounts of trust* und *functional accounts of trust*.

Trust as dependent on characteristics of the trusted beschreibt Vertrauen als Folge der besonderen Eigenschaften einer Person. So bestimmen beispielsweise Charaktermerkmale, ob

3. Theorie: Vertrauen

eine Person vertrauenswürdig ist und ihr Vertrauen entgegengebracht werden kann. (Hardin (2001a), S. 18ff)

Mit *Trust as a good* meint HARDIN ein Verständnis von Vertrauen als Ware. Diese Ware kann, wenn benötigt, produziert werden. Ebenso kann es gehandelt und zerstört werden oder in Form von sozialem Kapital für verschiedene Zwecke eingesetzt werden. (Hardin (2001a), S. 21ff)

Developmental accounts of trust gehen davon aus, dass die Fähigkeit, zu vertrauen und vertrauenswürdig zu sein, eine Folge von Erfahrung und Lerneffekten ist. Dabei können sowohl prägende Lebensereignisse als auch evolutionäre Mechanismen eine Rolle spielen. (Hardin (2001a), S. 24ff)

Functional accounts of trust zielen auf die Funktion von Vertrauen ab, die Komplexität, die sozialen Beziehungen innewohnt, reduzieren zu können. In der Regel kann ein Individuum nur schwer sicher abschätzen, wie sich der Partner innerhalb einer sozialen Beziehung verhalten wird. Es ist somit mit einer Unsicherheit bezüglich des Verhaltens des anderen und einer damit einhergehenden Verletzbarkeit seiner selbst konfrontiert, die sich aus der Komplexität sozialer Beziehungen ergibt. Vertrauen ist in der Lage diese Komplexität zu reduzieren und „increases tolerance for ambiguity“ (Hardin (2001a), S. 28). (Hardin (2001a), S. 28ff)

ENDRESS unterscheidet zwei prinzipielle Arten von Vertrauenskonzepten: *reflexives Vertrauen* und *fungierendes Vertrauen*. (Endress (2002), S. 47ff, S. 68ff, S. 90)

Reflexives Vertrauen beschreibt dabei Vertrauensansätze, die auf Risikoabschätzung beruhen. Solches Vertrauen ist auf spezifische soziale Situationen und das subjektive Risikobewusstsein der beteiligten Personen bezogen. Es beinhaltet die Kalkulation von Sanktionen gegen Vertrauensmissbrauch, die Bewertung des Kosten-Nutzen-Verhältnisses einer Vertrauensbeziehung sowie die explizite Kontrolle des Handelns der Vertrauenspartner. (Endress (2002), S. 68f)

Fungierendes Vertrauen meint Vertrauensansätze, die einen Modus von Vertrauen beschreiben, der implizit vorhanden ist und dabei jedwede Interaktion durchdringt. Vertrauen dieser Art bezieht sich auf die objektive Risikohaftigkeit von Vertrauenssituationen. Es ist im Sinne einer Präreflexivität, eine fraglose Form von Vertrauen, die eine Vertrautheit mit der Umwelt beschreibt. (Endress (2002), S. 68ff)

MISZTAL trennt in Vertrauenskonzeptionen, die Vertrauen als *property of individuals*, *property of social relationships* oder *property of social systems* begreifen.

Vertrauen als *property of individuals* ergibt sich als Folge individueller Persönlichkeitsmerkmale. Hierbei spielen insbesondere Empfindungen, Gefühle und individuelle Werte eine große Rolle. (Misztal (1996), S. 14)

Ansätze, die Vertrauen als *property of social relationships* verstehen, gehen davon aus, dass Vertrauen eine Eigenschaft von Kollektiven darstellt und im Rahmen bzw. durch die institutionelle Ordnung einer Gesellschaft wirksam wird. Vertrauen wird in dieser Lesart als soziale Ressource aufgefasst, die dazu dient gemeinschaftliche Ziele zu erreichen. (Misztal (1996), S. 14)

Vertrauen als *property of social systems* ist an die Handlungen und Einstellungen der Individuen innerhalb eines Systems gebunden. Vertrauen gilt hier als wertvolles öffentliches Gut, das durch das Verhalten der Mitglieder einer Gesellschaft aufrechterhalten wird. (Misztal (1996), S. 14f)

3. Theorie: Vertrauen

KUNZ vertritt die Auffassung, dass soziales Vertrauen der Förderung demokratischer Einstellungen und Verhaltensweisen dient, als soziales Koordinationsmedium sozialer Entwicklung und Integration wirkt sowie Transaktions- und Überwachungskosten senkt, was die Risikobereitschaft und die Innovationsbereitschaft stärkt. Er unterscheidet hierbei in *spezifische Vertrauensansätze* und *generalisierende Vertrauensansätze* (Kunz (2004), S. 202ff)

Unter *spezifischen Vertrauensansätzen* versteht er in erster Linie Vertrauensansätze, die Vertrauen als „rationale und fallbezogene Entscheidung“ (Kunz (2004), S. 204) betrachten. Die Entscheidung, Vertrauen zu schenken oder zu verweigern, orientiert sich dabei an der jeweiligen Situation, mit der man gerade konfrontiert ist, und ist abhängig von den Informationen, über die man hinsichtlich des Vertrauensnehmers und der Situation verfügt. Hierbei kann niemals eine Vollinformiertheit vorliegen, die Vertrauensvergabe ist somit immer mit einem Risiko verbunden.

Generalisierende Vertrauensansätze beziehen sich dagegen „auf die grundsätzliche Vertrauensbereitschaft eines Akteurs, unabhängig von den konkreten Ausprägungen einer bestimmten Situation“ (Kunz (2004), S. 204). Vertrauen ist in diesem Sinne die generalisierte Einstellung, dass den Mitmenschen eine grundsätzliche Vertrauenswürdigkeit innewohnt. Diese Einstellung ergibt sich aus während der Sozialisation erlernten kulturellen Regeln und Wertorientierungen.

STOLLE unterscheidet insgesamt vier Gruppen von Vertrauensansätzen: *Strategic or Rational Accounts of Trust*, *Identity- or Group-based Accounts of Trust*, *Moral Trust* und *Generalized Trust*. (Stolle (2002), S. 400ff)

Unter *Strategic or Rational Accounts of Trust* fallen solche Theorieansätze, die Vertrauen als Kosten-Nutzen-Rechnung bezüglich zukünftiger Kooperation verstehen. Vertrauen wird nach diesem Verständnis als Entscheidung zur Kooperation mit einer anderen Person betrachtet, die auf einer Abschätzung der Wahrscheinlichkeit beruht, ob diese andere Person das Vertrauen erwidern wird bzw. eine Gegenleistung für das erwiesene Vertrauen erbringen wird. Die Prognose des zukünftigen Verhaltens basiert dabei auf der Wahrnehmung der Performanz der fraglichen Person in früheren, möglicherweise ähnlich gelagerten Situationen. Um Vertrauen zu schenken, ist es daher notwendig bereits ein gewisses Maß an Informationen über den Vertrauensnehmer zu besitzen. Nachteil einer solchen Vertrauenskonzeption ist allerdings, dass sie nicht erklären kann, wie eine Vertrauensbeziehung zu Personen, über die nur wenige oder gar keine Informationen zur Verfügung stehen, etabliert werden kann. (Stolle (2002), S. 400f)

Identity- or Group-based Accounts of Trust basieren auf der Identifikation mit Gruppen und der Klassifizierung von Individuen. Annahme dieses Theorieansatzes ist es, dass „people trust those whom they feel close, whom they believe are similar to them, and with whom they are familiar“ (Stolle (2002), S. 401). Mit anderen Worten: Personen vertrauen solchen Personen, mit denen sie die gleiche Gruppenidentität vereint. Werden andere Personen als zur gleichen Gruppe zugehörig wie man selbst angenommen, so wird implizit davon ausgegangen, dass man mit ihnen die gleichen Werte und Normen teilt. Dies führt wiederum zu der Annahme, dass solche Personen in ähnlicher Art und Weise handeln würden wie man selbst. Gleichzeitig sorgt die Annahme einer gemeinsamen Gruppenidentität für die Perzeption einer gesteigerten Wahrscheinlichkeit von gruppengestützten Sanktionen bei einem Vertrauensmissbrauch. Die Folgen sind eine Reduktion des Kooperationsrisikos und eine damit einhergehende Erleichterung in eine Vertrauensbeziehung einzutreten. Das Konzept des sogenannten *depersonalized trust* erlaubt es dabei Vertrauen auf Gruppenmitglieder auszudehnen, die einem

3. Theorie: Vertrauen

persönlich nicht bekannt sind. (Brewer (1981), S. 356) Eine Erklärung für die Erweiterung von Vertrauen auf Personen, über die nur wenige oder gar keine Informationen bekannt sind bzw. die nicht als Mitglied unserer Gruppe identifiziert werden können, liefern auch diese Ansätze nicht.

Moral Trust basiert auf der Annahme, dass Personen grundsätzliche Werte miteinander teilen. Vertrauen ist innerhalb dieser Gruppe von Vertrauenskonzeptionen als grundsätzliche Einstellung des Menschen gegenüber anderen Menschen, im Sinne eines Glaubens an das Wohlwollen der anderen, zu verstehen. (Seligman (1997) S. 43) Solches Vertrauen ist unabhängig von Erfahrung und bleibt auch bei offensichtlichen Vertrauensbrüchen stabil. Unklar bleibt allerdings, warum verschiedene Personen unterschiedliche Vertrauenslevel aufweisen, wenn doch Vertrauen eine gegenüber äußeren Einflüssen robuste Charaktereigenschaft von Individuen und sogar Kollektiven darstellt.

Generalized trust bezieht sich im Speziellen auf Personen, die einem nicht persönlich bekannt sind. Es ist dabei gekennzeichnet durch „an abstract preparedness to trust others and to engage in actions with others“ (Stolle (2002), S. 403). *Generalized trust* überwindet die Notwendigkeit eine persönliche Beziehung zu jemandem zu haben, um ihm zu vertrauen. Allerdings scheint dabei der Kontext eine Rolle zu spielen.

NANNESTAD schließlich schlägt eine Klassifizierung der verschiedenen Vertrauensansätze vor, die sich auf zwei Ebenen bewegt. Auf der ersten Ebene unterscheidet er zwischen *rationalen Vertrauenskonzepten* und *norm-basierten Vertrauenskonzepten*. Die zweite Ebene gliedert er in *generalisiertes* und *partikulares Vertrauen*. (Nannestad (2008), S. 414f) Nach den *rationalen Vertrauenskonzepten* ergibt sich Vertrauen aus einer Abschätzung der zu erwartenden Konsequenzen einer Vertrauensvergabe. Ziel einer Vertrauensvergabe soll die Kontinuität einer Kooperationsbeziehung sein, wobei die Motivation des Vertrauenspartners, unseren Interessen zu entsprechen sowie seine Fähigkeit unseren Interessen zu dienen, eine große Rolle spielen. Zusätzlich wird Vertrauen dabei nur in Bezug auf eine bestimmte Sache oder einen bestimmten Sachverhalt wirksam. *Norm-basierte Vertrauenskonzepte* verstehen Vertrauen als sozialisations-induzierte grundsätzliche Einstellung gegenüber den Mitmenschen. Diese ist weitestgehend unabhängig von persönlicher Erfahrung und Risikokalkulation. Sowohl rationale als auch norm-basierte Vertrauenskonzepte lassen sich auf dem Kontinuum der zweiten Unterscheidungsebene klassifizieren. So bezeichnet *partikulares Vertrauen* ein Vertrauen, das auf eine bestimmte Person, hinsichtlich einer speziellen Angelegenheit gerichtet ist, wobei über diese Person Informationen bekannt sind. Dagegen bezeichnet *generalisiertes Vertrauen* ein Vertrauen gegenüber Fremden oder Vertrauen hinsichtlich eines nicht näher bestimmten Sachverhalts. (Nannestad (2008), S. 414f)

Betrachtet man die verschiedenen hier vorgestellten Systematisierungen des Vertrauensbegriffs, so ist eine Entwicklung des Verständnisses von Vertrauen zu beobachten, die in den Ansätzen von STOLLE und NANNESTAD konvergiert. Während JONES und HARDIN Vertrauen eher als Phänomen der Partikularebene betrachten und es vornehmlich in sozialen Beziehungen zwischen Individuen verorten, integrieren ENDRESS, MISZTAL und KUNZ zusätzlich die Generierung von Vertrauen auf der Ebene eines gesellschaftlichen Systems. Vertrauen ist dann nicht mehr nur Folge eines expliziten Aushandelns zwischen zwei oder mehr Personen, sondern kann auch implizit als Folge der Verfasstheit einer Gesellschaft existieren. Es erfolgt

3. Theorie: Vertrauen

somit eine Unterscheidung zwischen einer spezifischen Form des Vertrauens und einer generalisierten Form des Vertrauens.

Die Systematisierungen von STOLLE und NANNESTAD differenzieren diese erste Integration weiter aus. STOLLE unterscheidet Vertrauen entlang der Anfangsbedingungen seiner Entstehung. Vertrauen kann hiernach Folge von Kosten-Nutzen-Abwägungen, Gruppenidentifikation, Normen oder generalisierten Vertrauenseinstellungen sein. Dabei beinhalten ausschließlich Vertrauensansätze der letzteren Art eine Erklärung dafür, wie Vertrauen auf unbekannte andere bzw. Fremde erweitert werden kann. Eine Trennung zwischen *partikularem* und *generalisiertem Vertrauen* ergibt sich damit aus der Differenzierung der Anfangsbedingungen. NANNESTAD unterscheidet zwei Dimensionen des Vertrauens. Die erste Dimension betrifft dabei ebenso wie bei STOLLE die Anfangsbedingungen von Vertrauen und trennt in rationales und normen-geleitetes Vertrauen. Die zweite Dimension trägt dem Umstand Rechnung, dass Vertrauen sowohl partikularen als auch generalisierten Charakter besitzen kann. Beide Dimensionen greifen ineinander, sodass der Typus des rationalen Vertrauens vornehmlich mit *partikularem Vertrauen* zusammenfällt, während der Typus des normen-geleiteten Vertrauens vornehmlich mit *generalisiertem Vertrauen* zusammenfällt. (Nannestad (2008), S. 415)

Die dem folgenden Kapitel zugrundeliegende Systematisierung von Vertrauen soll nun aus einer Integration der Strukturierungen von STOLLE und NANNESTAD bestehen. Vertrauen wird hiernach entlang zweier Dimensionen betrachtet werden. Die erste Dimension spannt dabei ein Kontinuum der Anfangsbedingungen von Vertrauen auf. Die Eckpunkte dieses Kontinuum liefert die Konzeption der ersten Dimension von NANNESTAD, während eine Adaption von STOLLEs Ansatz den Zwischenraum besetzt. Es ergibt sich dann folgende Zusammensetzung für die erste Dimension: *Normative or moral trust, emotional or affective trust, identity or group-based trust, strategic or rational trust*. Die Reihenfolge der Vertrauensarten ist dabei keineswegs willkürlich gewählt, sondern verweist darauf, dass ausgehend vom Typus des normen-basierten Vertrauens über emotionales und identitäts-basiertes Vertrauen hin zu rationalem Vertrauen, die Relevanz von normativen Regelung für die Entstehung von Vertrauen abnimmt und dafür die individuelle und sachspezifische Aushandlung von Vertrauen an Bedeutung gewinnt.

Die Gliederung der Vertrauensarten wurde, verglichen mit STOLLE, an zwei Punkten verändert. Existierten bei STOLLE noch *moral trust* und *generalized trust*, wurden diese beiden hier zu *normative or moral trust* zusammengezogen. Dies ist damit zu begründen, dass auch *generalized trust* ein Vertrauen darstellt, welches auf allgemeingültigen Normen und Werten basiert. Es handelt sich hierbei um den Kern des hier zu entwickelnden Verständnisses von *generalisiertem Vertrauen* und gleichzeitig um den wesentlichen Parameter für die Distinktion zwischen *generalisiertem* und *partikularem Vertrauen*. Im Verlauf des nachfolgenden Kapitels wird die Argumentation hierzu im Detail ausgearbeitet werden.

Die zweite Veränderung im Hinblick auf STOLLE ist die Ergänzung von *emotional or affective trust*. Diese Vertrauenskategorie trägt dem Umstand Rechnung, dass Vertrauen oft eine affektive Komponente besitzt, die in den übrigen Vertrauensansätzen unterrepräsentiert ist. Da Ansätze, die von einem emotionsgeleiteten Vertrauen ausgehen eine weitere, von den anderen Perspektiven stark abweichende, Lesart des Vertrauensbegriffs liefern, werden sie hier als gesonderte Kategorie aufgeführt.

Die zweite Dimension stellt eine Adaption der zweiten Differenzierungsebene des Vertrauens bei NANNESTAD dar. Sie verläuft quer zur ersten Dimension und geht zunächst davon aus, dass soziales Vertrauen zwei verschiedene Qualitäten aufweisen kann: Vertrauen (*trust*) und

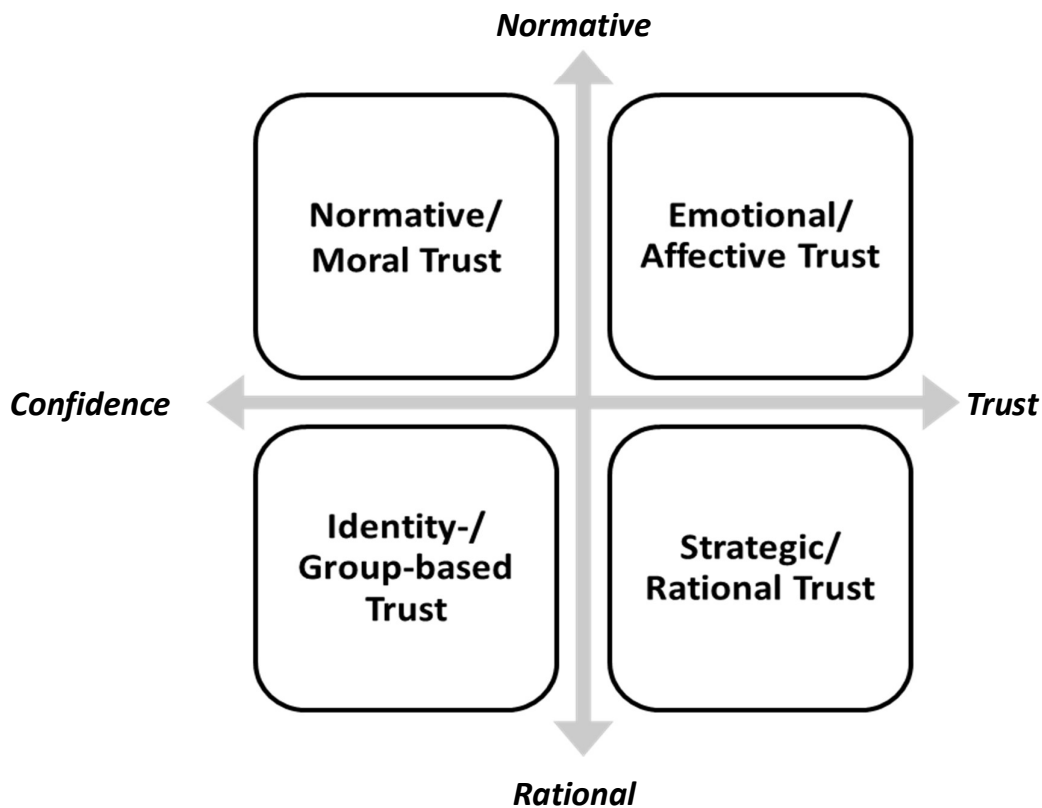
3. Theorie: Vertrauen

Zuversicht (*confidence*). Vertrauen (*trust*) beschreibt eine Vertrauensform, die einen spezifisch-reflexiven Charakter aufweist. Sie ist auf bestimmte Personen gerichtet und beinhaltet ein bewusstes Risiko. Zuversicht (*confidence*) beschreibt eine generelle Vertrauenseinstellung, auf deren Basis fraglos und selbstverständlich Vertrauen gewährt wird. Beide Vertrauensqualitäten konvergieren in einer jeweils charakteristischen Ausdruckform. Für Vertrauen (*trust*) stellt dies das sogenannte *partikulare Vertrauen* dar, welches auf bekannte Personen wegen einer bestimmten Angelegenheit oder eines bestimmten Gegenstandes bezogen ist. Für Zuversicht (*confidence*) ergibt sich das sogenannte *generalisierte Vertrauen*, dass vornehmlich auf Fremde bezogen ist und losgelöst von *face-to-face*-Interaktion bzw. persönlicher Bekanntschaft existiert.

Nachfolgend ist der Verlauf der ersten und zweiten Dimension sozialen Vertrauens zueinander grafisch dargestellt. Auf der vertikalen Achse befindet sich das Kontinuum der ersten Dimension, welches zwischen den Endpunkten *normative* und *rational* pendelt. Es ist dabei ausgehend von *Normative* oder *moral trust* über *Emotional* oder *affective trust*, *Identity*- oder *group-based trust* hinzu *Strategic* oder *rational trust* prinzipiell von einer Abnahme der Relevanz normativer Regelungen für die Entstehung von sozialem Vertrauen und einer Zunahme der Bedeutung der individuellen und sachspezifischen Aushandlung von sozialem Vertrauen auszugehen. Auf der horizontalen Achse ist die quer zur ersten Dimension verlaufende zweite Dimension sozialen Vertrauens zu finden. Die Pole des hier aufgespannten Kontinuums stellen die Kategorien Zuversicht (*confidence*) und Vertrauen (*trust*) bzw. *generalisiertes Vertrauen* und *partikulares Vertrauen* dar.⁹ Dabei enthält jede Kategorie der ersten Dimension sowohl Aspekte des *partikularen* wie des *generalisierten Vertrauens*. Allerdings sind diese je nach Kategorie mehr oder weniger stark ausgeprägt.

⁹ Auf die im Rahmen der zweiten Dimension Sozialen Vertrauens aufgespannte Differenz zwischen Zuversicht (*confidence*) und Vertrauen (*trust*) wird in einer nachfolgenden Veröffentlichung detailliert eingegangen werden. An dieser Stelle soll nur der Unterschiede zwischen den praktischen Manifestationen dieser beiden Vertrauenskategorien, Partikulares und Generalisiertes Vertrauen, thematisiert werden.

Abb. 3.1: Dimensionen sozialen Vertrauens



(Quelle: eigene Darstellung)

3.1.1 Ausdifferenzierung Sozialen Vertrauens

Die erste Dimension der Differenzierung von sozialem Vertrauen befasst sich mit der Unterscheidung normativer, affektbasierter, identitätsbasierter und rationaler Vertrauenskonzeptionen.

3.1.1.1 *Normative oder moral trust*

Normative Vertrauenskonzepte gehen davon aus, dass die Vergabe und Aufrechterhaltung von Vertrauen an die Befolgung allgemeingültiger und auch allgemein anerkannter Normen und Regeln gebunden ist. Vertrauen wirkt dabei ordnungstiftend und hat einen genuin sozialen Charakter, was bedeutet, dass es außerhalb von sozialen Beziehungen keine Bedeutung hat. Normative Vertrauensansätze nehmen an, dass Vertrauen weniger durch persönliche Erfahrungen beeinflusst ist, als dass es eine innerhalb einer Gesellschaft vererbte kulturelle Maxime darstellt. Vertrauen beschreibt hier demnach eine generelle Verhaltensdisposition, die jedem Mitglied einer Gesellschaft, in der ein Klima des Vertrauens vorherrscht, innewohnt und qua Sozialisation weitergegeben und reproduziert wird. Vertrauen wird zu einer moralischen bzw. kulturellen Ressource eines Gesellschaftsgefüges, wobei Vertrauen dann im Kern auf

3. Theorie: Vertrauen

gemeinsame und miteinander geteilte soziale Regeln gerichtet ist. Das wechselseitige Vertrauen in die Allgemeingültigkeit dieser gesellschaftlichen Normen und Regeln sowie die daran gebundene Annahme, dass diese Regeln und Normen auch von allen pflichtgemäß befolgt werden, können dann als Stifter von sozialer Ordnung wirken. Als innerhalb eines Gesellschaftsgefüges verankertes Vertrauen, ist normbasiertes Vertrauen dabei weitestgehend unempfindlich für Vertrauensbrüche bzw. –verletzungen.

Elemente dieses Aspektes finden sich in allen normativen Vertrauenskonzeptionen wieder. Am deutlichsten tritt er in den Konzeptionen von SZTOMPKA, BAIER, GOVIER, ERIKSON, BARBER, GARFINKEL und SELIGMAN zu Tage. SZTOMPKA beschreibt mit seiner *culture of trust* ein innerhalb einer Gesellschaft allgemeingültiges und alles durchdringendes Klima des Vertrauens, in dem Vertrauen zu einer typischen Leitlinie des Handelns wird. (Sztompka (1995), S. 259f; Sztompka (1996), S. 42f; Sztompka (1998), S. 21f; Sztompka (1999), S. 99ff)

ERIKSONs *Urvertrauen* wird bereits im Kindesalter angelegt und ist als ein Gebot, andere als vertrauenswürdig zu betrachten, zu verstehen. (Erikson (1970), S. 102f; Erikson (1974), S. 241f) BAIER spricht von einem *climate of trust*, nach dem im alltäglichen Handeln Vertrauen als selbstverständlich angenommen wird. Es wird dabei darauf vertraut, dass gesellschaftliche Rollen pflichtgemäß befolgt werden. (Baier (1986), S. 234f, S. 245f; Baier (1992), S. 147ff)

GOVIER beschreibt ein *basic trust*, das es in erster Linie möglich machen soll, mit Fremden zu interagieren, und in diesem Sinne eine Notwendigkeit für die Wechselbeziehungen in modernen differenzierten Gesellschaften darstellt, in denen ein Großteil der Interaktionspartner des täglichen Lebens einem persönlich unbekannt sind. Auch hier wird im Speziellen auf eine pflichtkonforme Rollenperformanz der einzelnen Gesellschaftsmitglieder vertraut. Aber auch Institutionen oder Kollektive können Gegenstand solchen Vertrauens werden. (Govier (1994), S. 238ff) BARBER ordnet Vertrauen als eine Erwartungshaltung gegenüber der Beständigkeit sozialer Ordnung ein. Als solches ist Vertrauen Grundvoraussetzung für gesellschaftliche Interaktion und reduziert gleichzeitig die Komplexität der sozialen Umwelt. Vertrauen in die Beständigkeit sozialer Ordnung verringert dabei die als legitim geltenden Optionen innerhalb der sozialen Interaktion. Bestimmte Handlungsweisen in der alltäglichen Interaktion können ausgeschlossen werden, weil sie nicht mit den Vorgaben bestimmter innerhalb des Gesellschaftsgefüges geltender sozialer Normen konform gehen. Entsprechend verringern sich auch die Möglichkeiten des Anschlusshandelns. (Barber (1983), S. 9)

GARFINKEL hält Vertrauen für eine Bedingung stabiler konzertierter Handlungen. Auch hier spielt wieder die wechselseitige Erwartung der Gesellschaftsmitglieder, dass spezifische gemeinsame Regeln befolgt werden, die entscheidende Rolle. Alltägliches Routinehandeln wird dabei erst durch das Vertrauen in die fraglose Gültigkeit bestimmter *basic rules* bzw. Hintergrunderwartungen möglich. Dabei entspricht das Einverständnis mit diesen Regeln und die Annahme, dass alle anderen diese Regeln in gleicher Weise befolgen wie man selbst, dem Vertrauen. (Garfinkel (1963), S. 187ff, S. 190ff, S. 212f; Garfinkel (1967), S. 35; Patzelt (1987), S. 125)

Auch SELIGMAN verortet Vertrauen als ein Wissen dessen, was erwartbar ist. So erlaubt das Vertrauen in die Existenz und Persistenz eines Systems normativer Rollenerwartungen eine wechselseitige Vorhersagbarkeit der Handlungen der Akteure. (Seligman (1997), S. 24f, S. 27f, S. 38f, S. 42; Seligman (1998), S. 391, S. 393f; Seligman (2000), S. 67, S. 72f)

In etwas abgeschwächter Form findet sich das obengenannte Charakteristikum normativer Vertrauenskonzeptionen auch bei SIMMEL, LUHMANN und GIDDENS. SIMMEL beschreibt mit seinem Verständnis von Vertrauen als einer Form induktiven Wissens eine Zuversicht gegenüber

3. Theorie: Vertrauen

der Konstanz der gesellschaftlichen bzw. natürlichen Ordnung. Im Fokus eines solchen Vertrauens steht dabei auch der gesamtgesellschaftliche Konsens über die Allgemeingültigkeit der innerhalb der sozialen Interaktion genutzten Kommunikationsträger. (Simmel (1989), S. 216)

LUHMANN rekurriert auf ein Systemvertrauen, wonach wir uns in einer uns selbstverständlich vertrauten Lebenswelt bewegen. Diese intersubjektiv geteilte Wahrnehmung von Selbstverständlichkeit und Kontinuität der sozialen Umwelt erlaubt eine Reduktion der Komplexität von Interaktionsbeziehungen in differenzierten Gesellschaftsgefügen. Vertrauen fungiert gleichermaßen als Funktionsgarant des Systems und ist ordnende Kraft innerhalb des Systems. Dabei ist dieses Systemvertrauen losgelöst von persönlichem Vertrauen und stellt die Kontinuität intersubjektiv-geteilter Wirklichkeit sicher. (Dieckmann (2006), S. 148f; Endress (2002), S. 32; Luhmann (1979), S. 7, S. 24, S. 43, S. 48f, S. 57f, S. 74; Luhmann (1987), S. 180f; Luhmann (1988), S. 96, S. 102f; Luhmann (2001), S. 147, S. 156; Luhmann (2009), S. 8, S. 26 S. 28, S. 30, S. 55, S. 61 S. 77, S. 99)

Auch GIDDENS verbindet Vertrauen mit einem Glauben an die Stabilität und Kontinuität der sozialen Umwelt und die Gültigkeit der innerhalb dieser Umwelt vollzogenen alltäglichen Routinehandlungen. Daneben spielt auch explizit das Vertrauen in die Funktion sogenannter *abstrakter Systeme* und deren Vertreter eine große Rolle. (Giddens (1992), S. 22, S. 27, S. 79f, S. 92; Giddens (1996a), S. 33ff, S. 40ff, S. 43ff, S. 47, S. 48f, S. 72f, S. 103, S. 107ff, S. 112, S. 113ff, S. 117ff, S. 128, S. 136ff; Giddens (1996b), S. 318f; Giddens (1997a), S. 138f; Giddens (1997b), S. 123)

Bei SCHWEER, WEBER/CARTER, MEYER/DAVIS/SCHOORMAN, ROSE-ACKERMAN und LEWIS/WEIGERT findet sich eine moralisch-normative Vertrauenskomponente, die eher auf der Individualebene wirksam wird und an unmittelbare Interaktionsbeziehungen von einzelnen sozialen Akteuren gebunden ist. SCHWEER verortet die normative Komponente des Vertrauens im einzelnen Individuum. So ist jede Person im Besitz einer individuell verschiedenen impliziten Vertrauenstheorie, die festlegt, welchen normativen Erwartungen ein potentieller Vertrauenspartner entsprechen muss, damit ihm Vertrauen gewährt wird. (Schweer (1997), S. 5ff)

WEBER/CARTER identifizieren Vertrauen als bedeutenden Faktor sozialer Beziehungen, der gleichzeitig aus diesen entsteht und sie aufrechterhält. Dabei spielt insbesondere das Vertrauen in die wechselseitige Befolgung der Beziehungsnormen eine große Rolle. Es wird darauf vertraut, dass der Vertrauenspartner in der Lage ist, sich in einen selbst hineinzusetzen und auf Basis dieses Perspektivenwechsels zu handeln. Die Fähigkeit zum Perspektivenwechsel stellt dabei sicher, dass ein Konsens bezüglich der die Beziehung begleitenden Normen gefunden werden kann. (Weber/Carter (1997), S. 25; Weber/Carter (1998), S. 9ff, S. 21; Weber/Carter (2003), S. 3)

Nach MEYER/DAVIS/SCHOORMAN ist Vertrauen u.a. von den Fähigkeiten des Vertrauten, seinem Wohlwollen dem Vertrauenden gegenüber und zuletzt von der moralischen Integrität des Vertrauten abhängig. (Mayer/Davis/Schoorman (1995), S. 717ff, S. 720)

Bei ROSE-ACKERMAN stellt die moralische Verpflichtung des Einzelnen zu Ehrlichkeit die Grundlage von Vertrauen dar. Auf der Ebene reziproken Vertrauens kommt zusätzlich noch die moralische Verpflichtung zu einer wechselseitigen Reflexion geteilter Werte und Normen hinzu. (Rose-Ackerman (2001), S. 526f, S. 539)

LEWIS/WEIGERT verankern Vertrauen in sozialen Beziehungen, wobei es als deren Attribut außerhalb solcher Beziehungen keine Bedeutung hat. In dieser von ihnen als holistisch beschriebenen Perspektive steht die reziproke Orientierung von Vertrauen auf der Basis geteilter interpretativer Annahmen über die Natur der Beziehung im Vordergrund. Die

3. Theorie: Vertrauen

normative Komponente dieser Vertrauenskonzeption zeigt sich dann infolge der Salienz der Vertrauenswürdigkeit des Vertrauenspartners. Ist ein gewisses Maß an Information über die Zuverlässigkeit des anderen erreicht, entsteht Vertrauen durch einen sogenannten *cognitive leap* und eine kollektiv-geteilte *norm of trust* übernimmt die weitere Gestaltung der Vertrauensbeziehung im Sinne einer Verselbständigung des wechselseitigen Vertrauens. (Lewis/Weigert (1985a), S. 968f, S. 970; Lewis/Weigert (1985b), S. 455f, S. 464)

Vertrauen stellt innerhalb der normativen Vertrauensansätze keineswegs ein einheitliches Konzept dar. Vielmehr wird Vertrauen innerhalb der Mehrzahl der hier vorgestellten normativen Vertrauenskonzeptionen je nach Theoretiker in verschiedene Ebenen untergliedert.

Eine verhältnismäßig basale Differenzierung des Vertrauensbegriffes findet sich bei LUHMANN, GIDDENS, SELIGMAN, MEYER/DAVIS/SCHOORMAN, SZTOMPKA. Die grundlegende Unterscheidung findet innerhalb dieser Ansätze zwischen den Ebenen *Vertrauen (trust)* und *Zuversicht (confidence)* statt. Ausgangspunkt beider Differenzierungsebenen ist jeweils der Umgang mit einer unsicheren Zukunft. In Bezug auf *Vertrauen (trust)* ist dies im Speziellen das Problem doppelter Kontingenz, welches jedweder Art von sozialer Beziehung und Vertrauensbeziehungen im Besonderen anhaftet. Während *Zuversicht (confidence)* sich auf Unsicherheiten bezieht, auf die keine aktive Einflussnahme möglich ist und die damit mit einer nicht näher bestimmten Gefahr eines Fehlschlages verbunden sind, bezieht sich *Vertrauen (trust)* auf unsichere Situationen, die mit einem Risiko der Enttäuschung verbunden sind, wobei man dieses Risiko bewusst eingeht bzw. ihm auch aus dem Weg gehen könnte. Während LUHMANN, MAYER/DAVIS/SCHOORMAN und SELIGMAN auf dieser Stufe der Differenzierung verharren, wird sie von GIDDENS und SZTOMPKA noch um die Begriffe *Glaube (faith)* und *Hoffnung (hope)* erweitert. *Glaube* bezeichnet dabei die Überzeugung von der Zuverlässigkeit von Personen und Dingen. *Hoffnung* bezieht sich dagegen vornehmlich auf natürliche Ereignisse und bezeichnet ein rational nicht begründbares Gefühl der Zuversicht. (Endress (2002), S. 31; Giddens (1992), S. 34, S. 92; Giddens (1996a), S. 40, S. 44ff, S. 48f, S. 51, S. 73, S. 117ff S. 124; Luhmann (1979), S. 29; Luhmann (1988), S. 97f, S. 99f; Luhmann (2001), S. 147f, S. 151f; Luhmann (2009), S. 36; Mayer/Davis/Schoorman (1995), S. 712ff, S. 715f, S. 720; Seligman (1997), S. 7f, S. 13, S. 20f, S. 24f, S. 27f, S. 39f, S. 42ff, S. 62ff, S. 69f, S. 70; Seligman (1998), S. 391ff S. 398f; Seligman (2000), S. 65ff, S. 68f, S. 72; Sztompka (1995), S. 255ff; Sztompka (1996), S. 38f S. 40ff Sztompka (1998), S. 19ff; Sztompka (1999), S. 18ff S. 24 S. 27f S. 30ff S. 41ff S. 53ff S. 60ff)

SIMMEL und BAIER schließen an die vorangegangene Systematisierung an. So unterscheidet SIMMEL zwei Differenzierungsebenen. Auf der ersten Ebene unterscheidet er ein persönliches, ein versachlichtes und ein universales Vertrauen. Auf der zweiten Ebene unterscheidet er in Vertrauen als Glauben, Vertrauen als Gefühl und Vertrauen als Wissen. Persönliches und versachtlichtes Vertrauen sowie Vertrauen als Wissen korrespondieren dabei mit der bereits zuvor aufgestellten Kategorie *Vertrauen (trust)* und beschreiben dementsprechend ein reflexives und auf die persönlichen Qualitäten eines Individuums bezogenes Vertrauen. Universales Vertrauen sowie Vertrauen als Glaube und Vertrauen als Gefühl entsprechen der Kategorie *Zuversicht (confidence)*. Allen dreien ist gemeinsam, dass sie ein Vertrauen beschreiben, das unreflektiert bleibt bzw. dem eine gewisse Fraglosigkeit innewohnt. BAIER unterscheidet lediglich zwischen einem Sich-Verlassen auf die Gewohnheiten eines anderen (=Verlass) und einem Sich-Verlassen auf das Wohlwollen des anderen (=Vertrauen). *Verlass* kommt dabei *Zuversicht (confidence)* gleich, während *Vertrauen* dem bereits beschriebenen

3. Theorie: Vertrauen

Vertrauen (trust) entspricht. (Baier (1986), S. 234f, S. 245f; Baier (1992), S. 147ff; Endress (2002), S. 14; Simmel (1989), S. 214, S. 216, S. 668f; Simmel (1992), S. 383ff, S. 393ff)

ROSE-ACKERMAN und GOVIER unterscheiden im Wesentlichen zwischen einem Vertrauen in Fremde und einem Vertrauen in bekannte Personen. ROSE-ACKERMAN unterscheidet zwischen *generalized trust*, wonach Vertrauen als generelle Einstellung zu verstehen ist, und *relational trust*, wonach Vertrauen einseitig oder wechselseitig ausgehandelt wird. GOVIER schließlich unterscheidet Vertrauen in Personen, mit denen man eine intime Beziehung pflegt, Vertrauen in vertraute Fremde und Vertrauen in vollkommen fremde Personen. (Govier (1994), S. 238ff; Rose-Ackermann (2001), S. 532ff, S. 536, S. 539f)

SCHWEER, WEBER/CARTER, LEWIS/WEIGERT wählen eine andere Art der Untergliederung des Vertrauens. Einig sind sich alle darin, dass Vertrauen eine kognitive Komponente besitzt, d.h. dass Informationen hinsichtlich der Vertrauenswürdigkeit des Vertrauten bekannt sein müssen bzw. man zum Perspektivenwechsel mit dem Vertrauten in der Lage sein muss. SCHWEER und LEWIS/WEIGERT stimmen weiter darin überein, dass Vertrauen zusätzlich eine affektiv-emotionale und eine behaviorale Komponente beinhaltet. Die affektive Komponente beschreibt dabei eine emotionale Involviertheit in die Vertrauensbeziehung und positive oder negative emotionale Verbundenheit mit dem Vertrauten. Die Verhaltenskomponente ist gleichzeitig Resultat der kognitiven und affektiven Komponente wie Verstärker der Vertrauensbereitschaft des Vertrauenden. So führen die Ausprägungen der kognitiven und affektiven Komponenten zu einem spezifischen Vertrauenshandeln, während die Anzeige von Vertrauen durch das Verhalten anderer sich positiv auf die eigene Vertrauensbereitschaft auswirkt. WEBER/CARTER ordnen Vertrauen neben der kognitiven Komponente noch eine moralische und eine soziale Komponente zu. Die moralische Komponente beschreibt dabei die Annahme, dass der Vertraute sich in seinem Verhalten an bestimmten Normen orientieren wird, während die soziale Komponente darauf verweist, dass Vertrauen aus sozialen Beziehungen entsteht und diese aufrechterhält. (Schweer (1997), S. 3, S. 7, S. 10; Lewis/Weigert (1985a), S. 970f, S. 972f; Lewis/Weigert (1985b), S. 464; Weber/Carter (1997), S. 25; Weber/Carter (1998), S. 9ff; Weber/Carter (2003), S. 2f)

ERIKSON unterscheidet zwischen Vertrauen in die Zuverlässigkeit anderer Personen sowie der Konstanz der Umwelt und Selbstvertrauen. Ausgangspunkt ist hierbei das sogenannte Urvertrauen, das die individuelle Ausstattung des Einzelnen mit Vertrauen darstellt. Während die erstgenannte Komponente dieses Urvertrauens auf die Umwelt gerichtet ist, sorgt das Selbstvertrauen dafür, dass man eine individuelle Vertrauenswürdigkeit entwickelt, die sowohl nach innen als auch nach außen wirksam wird. (Erikson (1970), S. 97, S. 100, S. 102f, S. 241ff)

BARBER und GARFINKEL beziehen sich in ihren Vertrauenskonzeptionen vornehmlich auf die Rolle von Vertrauen für die soziale Ordnung, die soziale Kohäsion und die soziale Kontrolle. (Barber (1983), S. 9, S. 10ff S. 19ff; Garfinkel (1963), S. 187ff, S. 212f, S. 216, S. 238; Garfinkel (1967), S. 35, S. 173)

Ein weiterer Aspekt normativer Vertrauensansätze ist die Diskussion der Relevanz von Vertrauen in modernen differenzierten Gesellschaften. Die Vielfalt der Rollen, die ein Individuum in modernen Gesellschaften ausübt, und die gleichzeitige Informationsfülle des Alltags macht es notwendig, bestimmte Kompetenzen auf andere gesellschaftliche Akteure auszulagern. Es ist dem Einzelnen schlichtweg nicht mehr möglich in allen Bereichen, die für das tägliche Leben relevant sind, erschöpfend informiert und qualifiziert zu sein. Deshalb wird es nun notwendig, Experten und deren Fachwissen Vertrauen entgegenzubringen, um diese Wissens- und Informationslücken zu schließen und dadurch das Funktionieren der Gesellschaft

aufrechtzuerhalten. Insbesondere SIMMEL, LUHMANN, GIDDENS, SZTOMPKA und SELIGMAN weisen in ihren Vertrauenskonzeptionen auf diesen Umstand hin. So beschreibt SIMMEL eine Versachlichung sozialer Beziehungen und Objektivierung von Kulturinhalten, die sich aus der Ausdifferenzierung moderner Gesellschaftsgefüge und der daraus resultierenden Zunahme gesellschaftlicher Komplexität ergeben. LUHMANN und GIDDENS betonen den fortschreitenden Bedeutungsverlust von persönlichem Vertrauen in modernen Gesellschaften und eine Verschiebung hin zu Vertrauen in das System. LUHMANN bezieht sich dabei u.a. darauf, dass nicht mehr nur uns unmittelbar bekannte und vertraute Personen Einfluss auf unser Leben ausüben, sondern auch Personen, zu denen kein unmittelbarer persönlicher Bezug besteht bzw. die uns mitunter nicht einmal bekannt sind, Teil unserer Lebenswelt sind. GIDDENS spricht in diesem Kontext von einer *Entbettung* sozialer Beziehungen aus lokalen Kontexten, sodass Gesellschaft zu einem Makrophänomen wird, das ohne das Vertrauen seiner Mitglieder nicht existieren kann. War Vertrauen zuvor noch auf ein lokales Gemeinwesen bezogen, in dem man sich wechselseitig bekannt war, ist es nun notwendig sich auf die Funktion sogenannter *symbolischer Zeichen* und *Expertensysteme* zu verlassen. SZTOMPKA trägt dieser Facette normativer Vertrauensansätze Rechnung, indem er auf die verschiedenen Gegenstände hinweist, auf die Vertrauen in modernen Gesellschaften gerichtet sein kann. Dies können u.a. konkrete Personen, soziale Kategorien oder soziale Rollen, aber auch soziale Gruppen, Organisationen oder Institutionen sein. (Dieckmann (2006), S. 148f; Endress (2002), S. 13, S. 32ff; (Giddens (1992), S. 21f S. 27 S. 102) (Giddens (1996a), S. 33ff, S. 40ff S. 72f, S. 103 S. 113ff S. 127ff S. 134ff, S. 136ff, S. 138ff; Giddens (1996b), S. 316ff, S. 319ff; Giddens (1997a), S. 138f Giddens (1997b), S. 116ff S. 123 S. 135ff; Lamla (2003), S. 108; Luhmann (1979), S. 7 S. 48ff, S. 50 S. 53, S. 54f S. 74 Luhmann (1988), S. 102; Luhmann (2001), S. 155ff; Luhmann (2009), S. 8 S. 30S. 60ff, S. 64 S. 69, S. 71f S. 78S. 99; Seligman (1997), S. 7f, S. 13, S. 20f S. 24f, S. 27f, S. 38f, S. 42 S. 54, S. 62 S. 70 Seligman (1998), S. 391ff; Seligman (2000), S. 65ff S. 72f; Simmel (1989), S. 215, S. 637, S. 668; Simmel (1992), S. 389, S. 393f; Sztompka (1995), S. 258ff Sztompka (1996), S. 42f Sztompka (1998), S. 21f Sztompka (1999), S. 102ff, S. 122ff, S. 125f, S. 128ff

3.1.1.2 *Emotional oder affective trust*

Emotions- bzw. affektbasierte Vertrauensansätze begreifen Vertrauen als Gefühl oder Empfindung, die auch körperlichen Ausdruck haben kann. Ausgangspunkt ist in der Regel eine affektive Reaktion auf die Wahrnehmung bzw. Annahme eines normativen Konsenses mit einem anderen Individuum. Diese Auffassung miteinander bestimmte Werte und Normen zu teilen, färbt die Weltwahrnehmung, sodass der andere vertrauenswürdig erscheint. Dabei ist diese Wahrnehmung relativ stabil gegenüber negativen Informationen.

Vertrauen beschreibt hier eine emotionale Verbindung zwischen den Vertrauenspartnern, die jegliche Weltwahrnehmung beeinflusst. Auf Seiten des Vertrauenden ist dabei ein affektgesteuerter Optimismus hinsichtlich des Wohlwollens und der Kompetenz des Vertrauten handlungsleitend. Der Vertraute erscheint dann auf Basis dieser Emotion als vertrauenswürdig, wobei diese Wahrnehmung gegenüber nachfolgenden eventuell negativen Informationen sehr robust ist. Sowohl die Vertrauenserwartung, nämlich sich auf den Vertrauten dahingehend verlassen zu können, dass dessen Handlungen zu einem positiven Ergebnis für einen selbst führen werden, und die anschließende Vertrauenshandlung, das Eingehen des Risikos sich

3. Theorie: Vertrauen

verletzbar zu machen durch die Handlungen des Vertrauten, sind hierdurch determiniert. Auf Seiten des Vertrauten soll sich durch das ihm entgegen gebrachte Vertrauen eine emotionale Ergriffenheit einstellen, die in einer Verbindlichkeit zwischen Vertrauendem und Vertrautem resultiert. Die Erfahrung, dass jemand auf ihn baut, soll dann das leitende Handlungsmotiv für den Vertrauten darstellen.

Bei DEDERICHS wird dieser Aspekt emotional-affektiven Vertrauens sichtbar, indem Vertrauen als emotionale Festlegungsstrategie beschrieben wird. Als emotionales Verhältnis dient Vertrauen dabei zur Überbrückung eines Informationsdefizites und die resultierende Verbindlichkeit zwischen Vertrauendem und Vertrautem bietet die Möglichkeit zur Etablierung eines Vertrauenskredites. (Dederichs (1997), S. 65f, S. 71ff, S. 75)

BARBALET deutet Zuversicht und Vertrauen als Empfindungen bzw. Gefühle zwischen den Vertrauenspartnern, die auch in Form von körperlichen Erfahrungen wirksam werden können. Vertrauen stellt dann die emotionale Basis von Kooperationsbeziehungen dar und ist gekennzeichnet durch die Zuversicht, dass die Handlungen des Vertrauten mit den eigenen Erwartungen korrespondieren. Wenn man vertraut, so herrscht nach BARBALET das Gefühl vor, sich auf den Vertrauten verlassen zu können. Dabei akzeptiert man die Abhängigkeit, in der man zum Vertrauten steht, ebenso wie die Abwesenheit von Informationen über die Verlässlichkeit des anderen, um einen ansonsten nicht realisierbaren Mehrwert zu erzeugen. (Barbalet (1993), S. 232; Barbalet (1996), S. 77; Barbalet (1998), S. 84S. 86 S. 96; Barbalet (2009), S. 367S. 368f S. 375)

JONES beschreibt den affektiven Aspekt des Vertrauens als optimistische Einstellung bezüglich dem Wohlwollen des Vertrauten gegenüber dem Vertrauenden seitens des Vertrauenden. Der Vertraute wird in der Folge durch eine *affective lens of trust* betrachtet, sodass jede seiner Handlungen nur noch vor dem Hintergrund des ihm erwiesenen Vertrauens interpretiert wird. Vertrauen bestätigt sich auf diese Weise ständig selbst und ist verhältnismäßig beständig gegenüber Ausdrücken von Vertrauensunwürdigkeit. Durch die Tatsache, dass auf ihn gezählt wird, entsteht beim Vertrauten eine emotionale Ergriffenheit, die ihn dazu anhält, sich dem erwiesenen Vertrauen als würdig zu erweisen. (Jones (1996), S. 4, S. 6ff, S. 11f, S. 16ff; Jones (2013), S. 5212)

MILLER zeichnet eine affektive Komponente des Vertrauens, deren Inhalt ein Gefühl ist, sich des Wohlwollens des Vertrauten und dessen Kompetenz hinsichtlich des Vertrauensgegenstandes sicher zu sein. Ebenso wie bei JONES stellt sich dann eine dem erwiesenen Vertrauen entsprechende Wahrnehmungsfärbung ein. (Miller (2000), S. 46f, S. 49ff; Miller (2002), S. 142f; Miller (2003), S. 18)

LAHNO begreift Vertrauen als Emotion. Ausgangspunkt dieses affektiven Vertrauens sind gemeinschaftlich anerkannte Normen, die determinieren, ob in einer spezifischen Situation Vertrauen angebracht ist oder nicht. Die Etablierung von Vertrauen verläuft dabei entlang eines automatischen normativen Konsenses über die jeweilige Situationsrahmung, an dessen Ende eine affektive Vertrauensreaktion steht. Die Folge ist auch hier eine Färbung der Weltwahrnehmung auf Basis dieser emotionalen Bindung zwischen den Vertrauenspartnern. Der Vertrauende nimmt den Vertrauten nun als eine Person wahr, mit der er gemeinsame Ziele und Werte zu teilen scheint, was wiederum diesen als grundlegend vertrauenswürdig erscheinen lässt. In der Folge werden vertrauensvolle Handlungsentscheidungen wahrscheinlicher. (Lahno (1999), S. 313; Lahno (2000), S. 88; Lahno (2001), S. 177, S. 182; Lahno (2002), S. 200ff, S. 210; Lahno (2004), S. 40f S. 43f; Lahno (2005), S. 99f, S. 101, S. 104ff; Lahno (2011), S. 675)

3. Theorie: Vertrauen

Ähnlich wie schon bei den normativen Vertrauensansätzen stellen auch die emotionsbasierten Vertrauenskonzeptionen kein homogenes Vertrauenskonzept dar. Entsprechend der Heterogenität des Vertrauensbegriffes ergeben sich auch hier je nach betrachtetem Theoretiker verschiedene Gliederungsebenen des Vertrauens.

Verhältnismäßig ähnlich in ihrer Aufschlüsselung des Vertrauens sind sich BARBALET, JONES und MILLER. BARBALET unterscheidet zunächst zwischen *Zuversicht* und *Vertrauen*, wobei Vertrauen eine besondere Form der *Zuversicht* darstellt, die im Speziellen auf Kooperation zwischen Individuen bezogen ist. *Zuversicht* stellt die emotionale Basis sozialen Handelns dar und kann auf die Überzeugung von der Fähigkeit eines anderen, etwas Bestimmtes zu tun, die Beurteilung der Sicherheit, mit der ein zukünftiges Ereignis eintritt, oder die Bereitschaft, auf Basis von Selbstvertrauen zu handeln, bezogen sein. *Zuversicht* wirkt dabei sowohl kognitiv als auch als Disposition, indem es Vorstellungen vom zukünftigen Ich und zukünftigen Ereignissen erzeugt sowie Handlungen auslöst, die auf Basis dieser Vorstellungen ausgeführt werden. Auf psychischer Ebene wird *Zuversicht* durch das Empfinden von Selbstvertrauen generiert, während es auf sozialer Ebene durch die Erfahrung von Akzeptanz und Anerkennung in sozialen Beziehungen entsteht. (Barbalet (1993), S. 231, S. 232ff, S. 235f; Barbalet (1996), S. 76ff, S. 81ff; Barbalet (1998), S. 84, S. 86ff; Barbalet (2009), S. 368f, S. 371)

JONES unterscheidet zunächst *Vertrauen* von *Verlass*. Dabei bezieht sich *Vertrauen* auf belebte Dinge, die über einen eigenen Willen verfügen, bzw. wenn das leitende Handlungsmotiv ein Auf-jemanden-zählen ist. Dagegen bezieht sich *Verlass* auf Maschinen oder ähnliche Dinge und kommt zum Einsatz, wenn Handlungsmotive wie Gewohnheit oder Angst etc. im Spiel sind, wenn keine Alternative besteht oder wenn keine Vorhersage über ein Ereignis möglich ist. *Vertrauen* gliedert sich dann in eine affektive und eine kognitive Komponente, wobei die affektive Komponente eine optimistische Einstellung gegenüber dem Wohlwollen des Vertrauten dem Vertrauenden gegenüber beschreibt und die kognitive Komponente, die aus dieser Einstellung resultierende Interpretation der Handlungen des Vertrauten seitens des Vertrauenden charakterisiert. (Jones (1996), S. 11f, S. 14f, S. 16ff; Jones (2013), S. 5212)

MILLER trennt zwischen *public trust* und *private trust*. *Public trust* beschreibt Vertrauen in Fremde, wobei davon ausgegangen wird, dass sich diese dem in sie gesetzten Vertrauen angemessen verhalten werden. *Private trust* spaltet sich in eine affektive und eine kognitive Komponente auf. Die affektive Komponente meint ein Gefühl, sich des Wohlwollens und der Kompetenz eines Vertrauten bezüglich eines spezifischen Vertrauensgegenstandes sicher zu sein. Dabei kann die Kompetenz als Fähigkeit des Vertrauten, etwas Bestimmtes tun zu können, oder als Fähigkeit des Vertrauten, die Erwartungen des Vertrauenden zu antizipieren, verstanden werden. Die kognitive Komponente beschreibt dann ähnlich wie bei JONES die Färbung der Wahrnehmung des Vertrauenden durch die affektive Komponente. (Miller (2000), S. 46f; Miller (2002), S. 142; Miller (2003), S. 17f, S. 20f)

LAHNO unterscheidet innerhalb des Vertrauens affektive, kognitive und behaviorale Elemente. Das affektive Element beschreibt die Verbundenheit zwischen Vertrauendem und Vertrautem, die sich als Folge der Annahme geteilter normativer Überzeugungen ergibt. Der Vertraute wird nun in einer spezifischen Art und Weise als vertrauenswürdig wahrgenommen. Gleichzeitig stellt das affektive Element die Grundlage der Vertrauensersparung und der Vertrauenshandlung dar. Das kognitive Element ist durch die Auffassung gekennzeichnet, dass der Vertraute ebenso wie der Vertrauende an die gleichen Normen gebunden ist und der Vertraute sich dementsprechend für den Vertrauenden vorteilhaft verhalten wird. Es besteht somit eine positive mentale Haltung gegenüber den potentiellen Handlungen des Vertrauten. Das behaviorale Element schließlich

drückt die eigentliche Vertrauenshandlung aus, in der man Vertrauen vergibt und sich verletzbar durch die Handlungen des Vertrauten macht. Es beschreibt die Abhängigkeit vom Handeln des Vertrauten, in die man sich begibt, wenn man Vertrauen schenkt. (Lahno (1999), S. 311, S. 313; Lahno (2000), S. 80ff; Lahno (2001), S. 171f; Lahno (2004), S. 30ff S. 32f S. 43f; Lahno (2005), S. 93ff S. 104ff; Lahno (2011), S. 675)

DEDERICHS unterscheidet verschiedene Formen des Vertrauens. Das emotionale Element, welches Vertrauen begleitet, ist dabei jeglicher dieser Formen immanent. (Dederichs (1997), S. 63f)

3.1.1.3 Identity oder group-based trust

Identitäts- bzw. gruppenbasiertes Vertrauen geht davon aus, dass für die Vertrauensvergabe eine wechselseitige Identifikation mit bestimmten Überzeugungen bzw. eine geteilte Gruppenmitgliedschaft leitend ist. Innerhalb dieser Vertrauensansätze wird angenommen, dass den Vertrauenspartnern bereits Informationen über diese Gemeinsamkeiten vorliegen und ihr Verhalten entsprechend danach ausrichten. Die Vertrauenspartner klassifizieren jemanden demnach entlang der Annahme als vertrauenswürdig, dass die Zugehörigkeit zur gleichen sozialen Kategorie die Werte und Normen determiniert, denen man sich verpflichtet fühlt. Angehörige der gleichen sozialen Gruppe sollten sich demnach einheitlicher Prinzipien verbunden fühlen, was die Handlungsweise dieser Personen für andere Gruppenmitglieder antizipierbar macht. Gleichzeitig sollen die Verbundenheit mit der sozialen Gruppe und die potentielle Sanktionierung von Fehlverhalten durch die Gruppe einen Vertrauensmissbrauch verhindern.

Solche Vertrauensansätze nehmen somit an, dass die wechselseitige Identifikation der Vertrauenspartner mit den Präferenzen des anderen bzw. die Identifikation mit der gleichen sozialen Gruppe leitend für die Vertrauensvergabe ist. Auf der Mikroebene stellt die am höchsten entwickelte Form des Vertrauens dann die vollkommene Internalisierung der Wünsche und Intentionen des Vertrauenspartners dar. Es wird dabei davon ausgegangen, dass ein wechselseitiges Verständnis füreinander vorherrscht und man ohne notwendige Überwachung für die Interessen des anderen eintritt. All dies geschieht im Bewusstsein dafür, welche Handlungen angebracht sind, um die Vertrauensbeziehung zu festigen. Auf der Makroebene ergibt sich Vertrauen aus einem *ingroup-outgroup*-Differential. Die Mitgliedschaft in sozialen Gruppen dient nun als Möglichkeit der Überbrückung von persönlicher Bekanntschaft. Die individuelle persönliche Identität einer Person wird dabei durch eine kollektive soziale Identität ersetzt, die ihrer Zugehörigkeit zu einer spezifischen sozialen Gruppe entspricht. Die Bewertung eines Individuums erfolgt dann anhand dieser sozialen Kategorie, wobei solche sozialen Gruppen favorisiert werden, denen man selbst angehört. Mitgliedern dieser sogenannten *ingroup* werden hinsichtlich ihrer Vertrauenswürdigkeit, Ehrlichkeit und Loyalität besser bewertet als Mitglieder der sogenannten *outgroup*. Eine geteilte bzw. gemeinsame Gruppenzugehörigkeit fungiert dann als Vertrauenskatalysator. Es wird dabei im Wesentlichen davon ausgegangen, dass Mitglieder der gleichen sozialen Gruppe sich den gleichen Normen und Werten verpflichtet fühlen und entsprechend ihrer Gruppenmitgliedschaft handeln werden. Als Versicherung dieses Handelns gilt gleichermaßen die jeweilige Gruppenmitgliedschaft, muss doch davon ausgegangen werden, dass Fehlverhalten gruppenintern sanktioniert wird.

3. Theorie: Vertrauen

SHAPIRO/SHEPPARD/CHERASKIN und LEWICKI/BUNKER betrachten identitäts-basiertes Vertrauen auf der Mikroebene und fokussieren dabei auf die wechselseitige Identifikation mit den Präferenzen des anderen bzw. die Internalisierung der Erwartungen des anderen. (Lewicki/Bunker (1995), S. 151ff, S. 156ff, S. 165ff; Lewicki/Bunker (1996), S. 122ff, S. 124f, S. 127f; Shapiro/Sheppard/Cheraskin (1992), S. 366; S. 372ff)

BREWER und KRAMER ET AL. beziehen sich verstärkt auf die Makroebene und haben die Identifikation mit spezifischen sozialen Gruppen als leitendes Vertrauensmotiv im Blick. Bei BREWER dient die Kenntnis der Mitgliedschaft in einer sozialen Gruppe als Mittel, um mangelnde persönliche Bekanntschaft auszugleichen. Innerhalb einer spezifischen sozialen Kategorie wird es auf diese Weise möglich, Opportunitäts- und Transaktionskosten von Kooperationsbeziehungen zu reduzieren. So wird angenommen, dass Individuen mit der gleichen Gruppenmitgliedschaft wie man selbst einem bestimmten sozialen Stereotyp entsprechen. Auf dieser Basis erfolgen eine Zuschreibung positiver Eigenschaften und die Annahme, dass man sich bei seinen Handlungen an die gleichen Prinzipien hält. (Brewer (1981), S. 351f, S. 355f; Brewer (2008), S. 215f, S. 217f; Maddux/Brewer (2005), S. 161; Yuki/Maddux/Brewer/Takemura (2005), S. 50)

KRAMER ET AL. gehen davon aus, dass die Mitgliedschaft in einer sozialen Gruppe dazu führt, dass psychische Transformationsprozesse einsetzen, in deren Folge Mitglieder dieser Gruppe in einer bestimmten Art und Weise wahrgenommen werden. Zwischen den Mitgliedern einer sozialen Gruppe reduziert sich auf diese Weise die soziale Distanz, sodass eine erhöhte Gleichheit untereinander wahrgenommen wird. Weitere Motive eine Vertrauensentscheidung zu treffen, sind außerdem die Annahmen, dass Vertrauen in gleicher Weise erwidert werden würde und dass Vertrauen innerhalb der Gruppe ein schützenswertes Kollektivgut darstellt. Vertrauen ergibt sich bei KRAMER ET AL. dann nicht aus der Verbundenheit der einzelnen Gruppenmitglieder, sondern aus der geteilten sozialen Identität. (Kramer/Brewer/Hanna (1996), S. 364ff, S. 366ff, S. 369, S. 373ff, S. 376f)

Bei TYLER ergibt sich Vertrauen u.a. durch die sozialen Bindungen zwischen den Mitgliedern einer sozialen Gruppe. Vertrauen entsteht aus Gruppennormen, die dazu animieren, die eigenen Interessen denen der Gruppe unterzuordnen und sich ganz den Regeln der Gruppe unterzuordnen. Wichtig für die Etablierung solchen Vertrauens sind dabei gemeinsame Werte und das wechselseitige Verständnis für die Handlungen des Vertrauenspartners. (Boeckmann/Tyler (2002), S. 2071; Tyler (2001), S. 288f, S. 302; Tyler (2003), S. 559, S. 562, S. 566f; Tyler/Degoey (1996), S. 344ff)

Auch identitäts- bzw. gruppen-basiertes Vertrauen stellt kein einheitliches Konstrukt dar, sondern wird je nach Theoretiker anders gefasst. Im Wesentlichen lassen sich Ansätze finden, die speziell auf die Mikroebene zugeschnitten sind, während andere Ansätze stärker die Makroebene betonen.

Eher auf der Mikroebene anzusiedeln sind dabei die Ansätze von SHAPIRO/SHEPPARD/CHERASKIN, LEWICKI/BUNKER und TYLER. Während die beiden Erstgenannten in ihren Konzeptionen sehr ähnlich sind bzw. aufeinander aufbauen, ist das Konzept von TYLER vergleichsweise eigenständig, dabei aber auch weniger ausdifferenziert. TYLER unterscheidet *rational* bzw. *calculative trust* von *social* bzw. *motive-based trust*. *Rational trust* beschreibt ein Vertrauen, bei dem es das Ziel ist, den Nutzen einer Vertrauensbeziehung zu maximieren. Hierzu wird sich bei der Vergabe von Vertrauen an Abschätzungen hinsichtlich des Verhaltens des Vertrauenspartners orientiert. *Social trust* wird den Mitgliedern einer spezifischen sozialen Gruppe entgegengebracht und ergibt sich aus der Identifikation mit dieser

3. Theorie: Vertrauen

sozialen Gruppe. Vertrauensleitend sind dabei gemeinsame Werte und wechselseitiges Verständnis für das Handeln des Vertrauenspartners. (Boeckmann/Tyler (2002), S. 207f; Brockner/Siegel/Daly/Tyler/Martin (1997), S. 558 Tyler (2001), S. 287, S. 301f; Tyler (2003), S. 559, S. 562, S. 566f; Tyler (2001), S. 288f, S. 302; Tyler/Degoey (1996), S. 332, S. 344ff)

SHAPIRO/SHEPPARD/CHERASKIN schlagen ein hierarchisches Vertrauensmodell vor. Die erste Stufe markiert dabei *deterrence-based trust*, welches davon ausgeht, dass Abschreckung durch Sanktionierung ein wirksamer Schutz vor Vertrauensmissbrauch ist. Auf der zweiten Ebene befindet sich *knowledge-based trust*, welches auf der Möglichkeit zur Voraussage der Vertrauenswürdigkeit eines anderen basiert. Auf der dritten Ebene folgt schließlich *identification-based trust*, welches sich aus der wechselseitigen Identifikation der Vertrauenspartner mit den Präferenzen des anderen ergibt. (Shapiro/Sheppard/Cheraskin (1992), S. 366ff; S. 369ff S. 372ff)

Das Modell von LEWICKI/BUNKER baut auf dem von SHAPIRO/SHEPPARD/CHERASKIN auf, geht allerdings davon aus, dass die einzelnen Vertrauensstufen nacheinander erklommen werden müssen und auch ein Rückfall auf die nächstniedrigere Stufe möglich ist. Dies kann beispielsweise bei einer besonders schweren Form des Vertrauensmissbrauchs geschehen. *Calculus-based trust* bildet dabei die erste Vertrauensebene und ist neben der Androhung von Sanktionen für Vertrauensverletzungen durch eine Belohnung der Vermeidung von Vertrauensverletzungen gekennzeichnet. *Knowledge-based trust*, auf der nächsten Vertrauensebene, basiert darauf aufgrund von Informationen über den Vertrauenspartner, dessen Verhalten antizipieren zu können. *Identification-based trust* auf der letzten Ebene fußt auf der vollkommenen Internalisierung der Erwartungen und Intentionen des Vertrauenspartners. (Lewicki/Bunker (1995), S. 145ff, S. 149ff, S. 151ff, S. 156ff, S. 163ff, S. 165ff; Lewicki/Bunker (1996), S. 119ff, S. 121f, S. 122ff, S. 124f, S. 126ff)

Die Ansätze von BREWER und KRAMER ET AL. zielen dagegen stärker auf die Etablierung von Vertrauen auf der Makroebene. So verlaufen nach BREWER Vertrauensentscheidungen entlang von Überlegungen hinsichtlich der Folgen, keine Vertrauensbeziehung einzugehen, hinsichtlich des Wertes des Nutzens, den man aus einer Vertrauensbeziehung erhält und hinsichtlich der Kosten, die einem aus dem Nutzen für den Vertrauten entstehen. Dabei ist in der Regel eine wechselseitige Bekanntschaft der Vertrauenspartner notwendig, die allerdings mit Hilfe von *depersonalized trust* überbrückt werden kann. In diesem Fall wird die individuelle Identität einer Person durch eine kollektive Identität ersetzt und Vertrauensentscheidungen auf Basis dieser kollektiven Identität getroffen. *Depersonalized trust* benötigt dabei ein wechselseitiges Wissen der Vertrauenspartner um eine geteilte Gruppenmitgliedschaft und basiert zusätzlich auf der Erwartung, dass sich der andere entsprechend dieser Gruppenmitgliedschaft verhalten wird. Grundlage des Vertrauens bildet somit die Überzeugung davon, dass innerhalb der fraglichen sozialen Gruppe bestimmte Normen gültig sind und Sanktionen durchgesetzt werden. Die Überbrückung mangelnder persönlicher Bekanntschaft bei der Vertrauensentscheidung kann allerdings auch auf dem Wege sogenannter *social identity processes* erfolgen. Hierbei ergibt sich Vertrauen als Folge eines automatischen Handelns auf der Basis von positiven Intragruppenerfahrungen (*heuristic decision making*), als Folge der Annahme der Kongruenz der eigenen Vorstellungen und der Intentionen anderer Gruppenmitglieder (*egoistic projection*) oder als Folge der vollkommenen Identifikation mit der sozialen Gruppe, wobei die eigenen Interessen den Interessen des Kollektivs untergeordnet werden (*social identification*). (Brewer (1981), S. 352f, S. 355f; Brewer (2008), S. 215ff, S. 220f, S. 223, S. 227; Kramer/Brewer/Hanna (1996), S. 360f; Maddux/Brewer (2005), S. 161; Yuki/Maddux/Brewer/Takemura (2005), S. 50)

KRAMER ET AL. gehen davon aus, dass die Identifikation mit einer sozialen Gruppe psychische Transformationsprozesse auslöst, die die individuellen Überzeugungen der Konsequenzen von Vertrauen und Misstrauen beeinflussen. Kognitive Transformationen lassen Mitglieder der eigenen sozialen Gruppe positiver erscheinen als Mitglieder anderer sozialer Gruppen. Motivationale Transformationen verstärken die Wahrnehmung von Gleichheit zwischen einem selbst und den anderen Gruppenmitgliedern. Affektive Transformationen betreffen den Spaß, der sich aus der Beteiligung an Gruppenprojekten ergibt. Dabei begleiten identitäts-basiertes Vertrauen verschiedene Motive. *Reciprocity-based trust* beschreibt die Vergabe von Vertrauen auf Basis der Annahmen, dass der Vertrauenspartner einem selbst ebenso Vertrauen schenken wird. *Elicitive trust behavior* nimmt an, dass der andere durch das Erweisen von Vertrauen der Vertrauenspartner dazu angeregt wird, ebenfalls zu vertrauen. *Compensatory trust* nimmt an, dass kollektives Vertrauen nur bewahrt werden kann, wenn man einer ausreichenden Menge an Personen Vertrauen entgegenbringt. *Moralistic trust* beschreibt ein Vertrauen, das auf ethischen Überzeugungen und Wertvorstellungen fußt. Um Vertrauen zu Fremden zu erklären, führen KRAMER ET AL. zusätzlich noch die Kategorie des *swift trust* ein, bei dem Vertrauen aus einem Handeln und Verhalten heraus entsteht, nach dem Vertrauen vorhanden sein müsste. Tatsächlich fehlen allerdings die klassischen Quellen des Vertrauens. Das vertrauensvolle Handeln ergibt sich dabei aus individuellen Prädispositionen, kategorialen Annahmen und impliziten Theorien. *Swift trust* ist dann auch eine Folge einer wechselseitig korrekt interpretierten Rahmung einer Situation. (Kramer/Brewer/Hanna (1996), S. 364ff, S. 366ff S. 369f, S. 371f, S. 373ff, S. 376f; Messick/Kramer (2001), S. 113f; Meyerson/Weick/Kramer (1996), S. 177f, S. 186ff, S. 191f)

3.1.1.4 *Strategic oder rational trust*

Rationale Vertrauensansätze begreifen Vertrauen als Kalkulation zukünftiger Kooperationsbeziehungen. Grundannahme ist, dass Vertrauen immer vom Risiko eines Vertrauensbruches begleitet wird. Ein Vertrauensbruch wird dabei mit einem Verlust für denjenigen gleichgesetzt, der vertraut hat. Vertrauen sollte demnach nur dann geschenkt werden, wenn zum einen aus der Vertrauensbeziehung zu erwartenden Gewinne die potentiellen Verluste aus einem Vertrauensmissbrauch übersteigen und wenn der Vertrauenspartner als hinreichend vertrauenswürdig angesehen wird. Vertrauen wird auf diese Weise zu einer Kosten-Nutzen-Relation. Um die Vertrauenswürdigkeit eines potentiellen Vertrauenspartners bzw. die Wahrscheinlichkeit eines Vertrauensmissbrauchs beim Eingang einer Vertrauensbeziehung mit dem Vertrauenspartner zu bestimmen, wird auf dessen Reputation zurückgegriffen oder dessen Performanz bei früheren Interaktionen zu Rate gezogen.

Rationale bzw. strategische Vertrauensmodelle nehmen an, dass die Vergabe von Vertrauen entlang einer rationalen Kosten-Nutzen-Abwägung verläuft, in deren Fokus die Verwirklichung der eigenen Interessen steht. Um Vertrauen zu gewähren, muss dabei eine Vertrauensschwelle erreicht werden, die sich aus dem Verhältnis der angenommenen Vertrauenswürdigkeit hinsichtlich des Vertrauten und der Höhe eines potentiellen Verlustes bzw. Gewinnes aus der Vertrauensbeziehung errechnet. Vertrauen ist dann lohnenswert, wenn die Vorteile aus der Vertrauensbeziehung die Risiken überwiegen bzw. die Eintrittswahrscheinlichkeit eines vorteilhaften Ereignisses größer ist als die eines nachteiligen Ereignisses. Grundlage sind dabei

3. Theorie: Vertrauen

auf der einen Seite die Risikobereitschaft und die Enttäuschungstoleranz des Vertrauenden sowie Informationen über die Reputation der Vertrauten und die daraus resultierenden Erwartungen hinsichtlich des potentiellen zukünftigen Verhaltens des Vertrauten. Dabei können die potentiellen Gewinne und Verluste einer Vertrauensbeziehung meist recht gut bestimmt werden, während hinsichtlich der Vertrauenswürdigkeit eines Vertrauten ein mehr oder weniger großes Informationsdefizit bestehen kann. Die Richtigkeit einer Entscheidung hängt davon ab, wie gut die Wahrscheinlichkeit vertrauenswürdigen Verhaltens antizipiert wird. Allerdings besteht auch immer das Risiko einer Unter- bzw. Überschätzung. Vertrauen bleibt in diesem Sinne eine riskante Vorleistung seitens des Vertrauenden, da sich der Vertraute immer noch als vertrauensunwürdig entpuppen kann. Es ist charakterisiert durch eine positiv gelagerte Voreingenommenheit auf der Basis unvollständiger Informationen hinsichtlich der Intentionen des Vertrauten, die Erwartung eines Wohlwollens seitens des Vertrauten, die Erwartung freundlichen Verhaltens in einer Situation, die durch Unsicherheit gekennzeichnet ist und die Erwartung, dass der Vertraute ein für den Vertrauenden vorteilhaftes Verhalten ausführen wird.

In Reinform findet sich dieser Aspekt rationalen Vertrauens bei DEUTSCH und COLEMAN. DEUTSCH beschreibt die Vertrauensentscheidung als Abwägung der Eintrittswahrscheinlichkeiten eines positiven Ergebnisses (Va+) und eines negativen Ergebnisses (Va-). Dabei ist der Ausgang einer Vertrauenssituation nicht klar determiniert, sondern beinhaltet eine Ungewissheit. (Deutsch (1960a), S. 124, S. 125; Deutsch (1962), S. 303f; Deutsch (1976), S. 136ff)

Für COLEMAN stellt die Vertrauensvergabe durch den Vertrauenden eine Kosten-Nutzen-Rechnung dar. Die Entscheidung zu vertrauen, basiert dabei auf dem Wissen um die Eintrittswahrscheinlichkeit eines bestimmten zukünftigen Verhaltens des Vertrauten. Es werden dabei der potentielle Gewinn wie Verlust, die aus der Vertrauensbeziehung erwachsen können, miteinander verrechnet und mit der Vertrauenswürdigkeit des Vertrauten in Beziehung gesetzt. Wird ein bestimmter Schwellenwert erreicht, so gilt die Vertrauensbeziehung als lohnenswert. Dabei liegen in der Regel recht genaue Informationen hinsichtlich der potentiellen Gewinne und Verluste vor, während hinsichtlich der Vertrauenswürdigkeit der Vertrauten normalerweise ein Informationsdefizit vorliegt, da diese nicht absolut genau bestimmt werden, sondern allenfalls abgeschätzt werden können. (Coleman (1982), S. 282f, S. 283ff; Coleman (1990), S. 104; Coleman (1991), S. 121f, S. 124ff, S. 126ff, S. 129ff, S. 137ff; Coleman (1994), S. 105; Endress (2002), S. 36; Junge (1998), S. 39ff)

Bei ZIEGLER, DASGUPTA, HARDIN, ROTTER, YAMAGISHI und GAMBETTA stehen die mit Vertrauen verbundenen Erwartungen seitens des Vertrauenden an den Vertrauten im Vordergrund. Vertrauen wird auch hier entlang von Informationen über den Vertrauten und auf diesen Informationen basierenden Abschätzungen hinsichtlich des potentiellen zukünftigen Verhaltens des Vertrauten gewährt. ZIEGLER versteht Vertrauen als riskante Vorleistung des Vertrauenden gegenüber dem Vertrauten, die an die Erwartung vertrauenswürdigen Verhaltens gebunden ist. Eine Vertrauensentscheidung wird dabei aufgrund der sozialen Interaktionen begleitenden doppelten Kontingenz immer auf Basis einer begrenzten Rationalität getroffen. (Ziegler (1997), S. 241ff, S. 245f, 252)

Bei DASGUPTA ist Vertrauen eine Erwartung hinsichtlich der Handlungen anderer, wenn diese Handlungen Auswirkungen auf die eigenen Handlungen haben werden und die Handlungen der anderen nicht im Vorhinein beobachtet werden können. (Dasgupta (1988), S. 51)

Laut HARDIN ergibt sich Vertrauen als Annahme über die Vertrauenswürdigkeit des Vertrauten auf Basis des Wissens, das man über diesen Vertrauten besitzt. Dabei kann sich Vertrauen nur

3. Theorie: Vertrauen

in fortlaufenden Beziehungen mit wiederholter und nutzenbringender Kooperation entwickelt werden, da nur in solchen Beziehungen die notwendigen Informationen über den Vertrauten generiert werden können. Vertrauen beschreibt dann die Erwartung, dass der Vertraute bestimmte für den Vertrauenden vorteilhafte Dinge tun wird sowie motiviert und befähigt ist, diese Dinge zu tun. (Hardin (1998), S. 11; Hardin (2001b), S. 295, S. 297; Hardin (2002), S. 3ff, S. 7S. 9f, S. 13ff; Hardin (2004), S. 6f; Hardin (2006), S. 17ff, S. 31)

ROTTER begreift Vertrauen als Erwartung, dass man sich auf die Versprechen eines anderen verlassen kann. Diese Erwartung setzt sich dabei aus spezifischen situationsgebundenen Komponenten und generalisierten Komponenten zusammen. Erstere sind nur in der konkreten Vertrauenssituation anwendbar, während letztere verallgemeinerte Erfahrungen aus vorangegangenen Vertrauenssituationen darstellen. (Rotter (1967), S. 651 S. 653; Rotter (1971), S. 445; Rotter (1980), S. 1f; Rotter (1981), S. 23)

YAMAGISHI geht davon aus, dass Vertrauen einen Glauben an das Wohlwollen eines Vertrauten darstellt, wobei hinsichtlich dieses Wohlwollens eine gewisse Unsicherheit besteht. Die Annahme des Wohlwollens fußt dabei auf der Wahrnehmung der Charaktereigenschaften des Vertrauten. (Yamagishi (2011), S. 23ff, S. 27, S. 31f; Yamagishi/Yamagishi (1994), S. 131f, S. 136)

Für GAMBETTA bedeutet Vertrauen, mit einer bestimmten Sicherheit davon überzeugt zu sein, dass der Vertraute kooperieren wird. Wann diese Vertrauensschwelle erreicht ist, hängt auf der einen Seite von den individuellen Veranlagungen des Vertrauenden und den objektiven Bedingungen der jeweiligen Vertrauenssituation ab. Auf der anderen Seite wird aber auch die Vertrauenswürdigkeit des Vertrauten berücksichtigt. Diese wird dabei sowohl anhand subtiler als auch manifester Zeichen abgeleitet. (Gambetta (2001), S. 211ff; Bacharach/Gambetta (2001), S. 148, S. 154f, S. 168ff)

Ebenso wie die vorangegangenen Vertrauenskategorien entwickeln auch die rationalen Vertrauensansätze kein homogenes Vertrauenskonzept und lassen sich grob in konzeptionelle Gruppen unterteilen. YAMAGISHI, ROTTER und DEUTSCH unterscheiden dabei innerhalb ihrer Vertrauensansätze in mindestens zwei Ebenen. Während YAMAGISHI und ROTTER im Wesentlichen in eine spezifische und eine generalisierte Ebene des Vertrauens unterscheiden, beschreibt DEUTSCH die Vertrauensentscheidung als *ambiguous path* und unterscheidet Vertrauensentscheidungen von Risikoentscheidungen. Die Beschreibung von Vertrauen als *ambiguous path* trägt dabei dem Umstand Rechnung, dass im Zuge von Vertrauen ein für den Vertrauenden vorteilhaftes oder nachteiliges Ergebnis resultieren kann, da er sich vom kontingenten Verhalten eines anderen Individuums abhängig macht. Die Gewährung von Vertrauen erfolgt nach DEUTSCH auf der Basis der Zuversicht, dass eher das vorteilhafte als das nachteilige Ereignis eintreten wird, wobei bei einer Vertrauensentscheidung die Eintrittswahrscheinlichkeit einer negativen Konsequenz größer ist als die einer positiven Konsequenz. Im Gegensatz dazu ist bei einer Risikoentscheidung die Eintrittswahrscheinlichkeit einer positiven Konsequenz höher als die einer negativen Konsequenz. Dies liegt darin begründet, dass bei einer Risikoentscheidung bereits einige Informationen hinsichtlich des Ausgangs einer Situation vorliegen, während bei einer Vertrauensentscheidung keine solchen Informationen verfügbar sind. (Deutsch (1960a), S. 124f; Deutsch (1962), S. 303f; Deutsch (1976), S. 136ff)

ROTTER versteht Vertrauen als eine generalisierte Erwartungshaltung, die sich von Individuum zu Individuum unterscheidet. Vertrauen beinhaltet dabei die Erwartung, dass man sich auf die Versprechen eines anderen verlassen kann. Die individuellen Unterschiede des Vertrauens

3. Theorie: Vertrauen

ergeben sich aus positiven und negativen Erfahrungen mit den Versprechen anderer, wobei sich Vertrauen aus der Verknüpfung zweier Erwartungsebenen ergibt. Spezifische Erwartungen stellen Erfahrungen und Begebenheiten dar, die an eine bestimmte Situation gebunden sind, während generalisierte Erwartungen aus Verallgemeinerungen von Gemeinsamkeiten verschiedener Situationen resultieren. (Rotter (1967), S. 653; Rotter (1971), S. 445; Rotter (1980), S. 2; Rotter (1981), S. 23)

YAMAGISHI entwickelt ein vergleichsweise stark ausdifferenziertes Vertrauenskonzept. Auf der ersten Differenzierungsebene unterscheidet er in *expectations of natural order* und *expectations of moral order*. Erstere beziehen sich auf die Annahme der Beständigkeit von Naturphänomenen, während letztere sich auf die Beständigkeit der sozialen Ordnung beziehen. Innerhalb der *expectations of moral order* unterscheidet er auf einer zweiten Ebene in *expectations of competence*, die die Zuversicht beschreiben, dass jemand auch in der Lage ist, das umzusetzen, was er versprochen hat zu tun, und *expectations of intentions*, die sich auf die Absicht eines anderen beziehen, etwas Versprochenes auch in die Tat umzusetzen. Die *expectations of intentions* differenziert er weiter aus in Vertrauen (*trust*) und Gewissheit (*assurance*). Dabei nimmt Vertrauen ein Wohlwollen seitens des Vertrauten an, das auf den Vertrauenden gerichtet ist, wobei sich diese Annahme aus den Charaktereigenschaften und Intentionen des Vertrauten ableitet. Vertrauen stellt darüber hinaus eine Voreingenommenheit seitens des Vertrauenden gegenüber dem Vertrauten dar, die auf die Verarbeitung unvollständiger Informationen hinsichtlich der Intentionen des Vertrauten wirkt. Vertrauen ist dann die Erwartung freundlichen Verhaltens seitens eines Vertrauten in einer sozialen Situation, in der Ungewissheit bezüglich dessen Handelns vorherrscht, die sich aus Überzeugungen hinsichtlich der Dispositionen des Vertrauten ergibt. Vertrauensbeziehungen sind somit dadurch gekennzeichnet, dass sie eine subjektive Reduktion sozialer Unsicherheit herbeiführen, objektiv gesehen allerdings das Risiko einer Enttäuschung weiterhin manifest bleibt. Gewissheit stellt dagegen eine Erwartung freundlichen Verhaltens dar, die sich aus anderen Motiven als Wohlwollen ergibt. Gewissheit basiert auf dem Wissen um die eine Beziehung umgebende Anreizstruktur, sodass die Annahme von freundlichem Verhalten seitens eines Interaktionspartners eine Folge der Überzeugung ist, dass solches Verhalten im Selbstinteresse des Interaktionspartners liegt. Gewissheit ergibt sich somit aus der Einschätzung, dass es für den Interaktionspartner keinen Anreiz gibt, den anderen auszunutzen. Entsprechend enthalten von Gewissheit geprägte Beziehungen auch keinerlei soziale Unsicherheit. Mit Gewissheit verknüpft sind sogenannte Verpflichtungsbeziehungen innerhalb derer die wiederholte Interaktion mit festen Partnern zu einer objektiven Reduktion sozialer Unsicherheit führt. Verpflichtungsbeziehungen reduzieren dabei zwar die Transaktionskosten, da sich das Risiko eines Vertrauensmissbrauchs bei wiederkehrender Interaktion mit festen Partnern verringert, gleichzeitig sorgt dieser Umstand aber auch für eine Erhöhung der Opportunitätskosten, da die feste Bindung an einen bestimmten Partner die Möglichkeiten zu anderen möglicherweise vorteilhafteren Verbindungen vermindert. Innerhalb von Vertrauen unterscheidet YAMAGISHI weitere Ebenen. Auf der ersten Ebene differenziert er hier in *information-based* bzw. *knowledge-based trust* und *generalized trust*. *Generalized trust* beschreibt dabei ein Vertrauen, das sich auf Individuen bezieht, über die keine spezifischen Informationen zugänglich sind, und ist als ein Glaube daran zu verstehen, dass freundliches Verhalten ein Aspekt der menschlichen Natur ist. *Information-based* bzw. *knowledge-based trust* beschreibt dagegen ein Vertrauen, das auf Informationen basiert, die sich aus einer längerfristigen Interaktionshistorie mit dem Vertrauten ergeben. YAMAGISHI differenziert diese Vertrauensform weiter in *relational* und

3. Theorie: Vertrauen

character-based trust aus. Ersteres betrifft Vertrauen, das sich aus Informationen über die Empfindungen und Einstellungen des Vertrauten gegenüber dem Vertrauenden ergibt, während Letzteres Vertrauen beschreibt, das sich aus Informationen über den Charakter des Vertrauten speist. *Character-based trust* wird außerdem noch in *personal trust* und *category-based trust* unterschieden. Dabei beschreibt *personal trust* Vertrauen, das sich aus Informationen über den Charakter eines bestimmten Individuums ergibt, und *category-based trust*, Vertrauen, das sich aus Informationen über die Zugehörigkeit eines Individuums zu einer spezifischen sozialen Gruppe ergibt. (Yamagishi (2011), S. 22f, S. 23ff, S. 25ff, S. 28ff, S. 32, S. 33ff, S. 31f; Yamagishi/Kikuchi/Kosugi (1999), S. 156f; Yamagishi/Yamagishi (1994), S. 131f, S. 136, S. 139)

Die Ansätze von GAMBETTA, DASGUPTA und ZIEGLER stellen das Vertrauensproblem in den Mittelpunkt ihrer Konzeptionen, welches im Wesentlichen darin besteht, dass Vertrauen eine Strategie zur Überbrückung von Ungewissheit innerhalb von sozialer Interaktion darstellt, dabei aber ein nicht unerhebliches Informationsdefizit hinsichtlich der wahren Intentionen des Vertrauten hinnimmt. GAMBETTA unterscheidet ein Vertrauensproblem erster Ordnung (Kann ich Person X vertrauen?) und eines zweiter Ordnung (Kann den Zeichen für die Vertrauenswürdigkeit einer Person getraut werden?). Für die Vertrauensentscheidung stellt Letzteres das relevantere Problem dar. Die Zeichen für Vertrauenswürdigkeit teilt GAMBETTA in zwei Arten ein, *krypta* und *manifesta*. Dabei stellen *krypta* solche Vertrauenszeichen dar, die keiner direkten Beobachtung zugänglich sind und sich aus der Kombination von Hintergrundwissen und Beobachtungen ableiten lassen. *Manifesta* stellen dagegen Vertrauenszeichen, die sich direkt aus dem Verhalten des potentiellen Vertrauten ergeben. (Gambetta (2001), S. 211ff; Bacharach/Gambetta (2001), S. 148, S. 154f, S. 168ff)

DASGUPTA befasst sich im Rahmen des Vertrauensproblems mit dem Umstand, dass nicht alle Menschen vertrauenswürdig sind, sondern sich oft Diskrepanzen zwischen individuellen und moralischen Werten ergeben. Seiner Ansicht nach ist Vertrauen bzw. Vertrauenswürdigkeit sowohl an die Reputation eines potentiellen Vertrauten sowie situationsspezifische Gelegenheiten gebunden. Vertrauen auf der Basis von Reputation ergibt sich aus früheren Handlungen und den damals begleitenden Handlungsmotiven und dem kulturellen bzw. gesellschaftlichen Hintergrund des potentiellen Vertrauten. Daneben spielen situationsgebundene Interessen der beteiligten Vertrauenspartner eine Rolle. (Dasgupta (1988), S. 53f)

ZIEGLER umreißt das Vertrauensproblem anhand von fünf Punkten: (1) Es soll sich nach Möglichkeit der jeweils andere als vertrauenswürdig erweisen. (2) Es besteht das Risiko eines Schadens aufgrund eines Vertrauensbruches. (3) Der Vertraute hat in der Regel einen Anreiz zum Vertrauensmissbrauch. (4) Vorzugsweise sollte sich wechselseitiges Vertrauen ergeben. (5) Der Vertrauende muss in der Regel den ersten irreversiblen Schritt zur Etablierung einer Vertrauensbeziehung machen. (Ziegler (1997), S. 242)

COLEMAN und HARDIN unterscheiden verschiedene Vertrauenskonstellationen und differenzieren dabei neben einseitigen und wechselseitigen Vertrauensbeziehungen auch zwischen Mikro- und Makroebene. Bei COLEMAN besteht der Prototyp einer Vertrauenskonstellation auf der Mikroebene aus einem sogenannten Treugeber und einem sogenannten Treuhänder, wobei Ersterer Letzterem die Kontrolle über in seinem Besitz befindliche Ressourcen übergibt, obwohl er keine gesicherte Aussage über dessen zukünftiges Handeln machen kann. Diese grundlegende Form der Vertrauensbeziehung kann auch als einseitiges Vertrauen beschrieben werden und wird im Folgenden von COLEMAN um sogenannte *Vertrauensintermediäre* erweitert. Der Berater fungiert als Kontaktvermittler

zwischen einem Treugeber und einem Treuhänder, indem er die Vertrauenswürdigkeit des Treuhänders bezeugt. Der Bürge fungiert als Mittler zwischen Treugeber und Treuhänder, indem er die Ressourcen des Treugebers an den Treuhänder weiterreicht und im Falle eines Vertrauensbruchs für den entstandenen Schaden haftet. Der Unternehmer vermittelt Treugeber an passende Treuhänder, sodass dem Treugeber ein möglichst hoher Nutzen erwächst. Neben den Vertrauensintermediären, verweist COLEMAN auf das sogenannte Drittparteien-Vertrauen, bei dem eine dritte passive Partei für die sich im Spiel befindenden Ressourcen garantiert. Eine wechselseitige Vertrauensbeziehung stellt sich dann ein, wenn beide Vertrauenspartner sowohl die Rolle des Treugebers als auch die Rolle des Treuhänders innehaben. Auf der Makroebene existieren nach COLEMAN im Wesentlichen zwei Vertrauenskonstellationen, die der *Gemeinschaft* und eine, die sich als ein *System beratender Intermediäre* darstellt. Eine Gemeinschaft stellt dabei ein System wechselseitigen Vertrauens dar. Innerhalb eines Systems beratender Intermediäre setzt sich Vertrauen gleich einem Schneeballsystem durch Weiterempfehlung eines Treuhänders fort, sodass am Ende alle Beteiligten einem bestimmten Treuhänder Vertrauen entgegenbringen. (Coleman (1982), S. 138ff, S. 282f, S. 283ff, S. 285ff, S. 288f, S. 290f, S. 293ff, S. 295ff; Coleman (1990), S. 178; Coleman (1991), S. 121f, S. 124ff, S. 127ff, S. 129ff, S. 137ff, S. 140ff, S. 228ff, S. 232, S. 233ff, S. 239ff, S. 242ff, S. 244ff, S. 354ff; Coleman (1994), S. 105 Endress (2002), S. 36; Junge (1998), S. 39ff, S. 43ff, S. 46ff)

HARDIN betrachtet Vertrauen als *encapsulated interest*, bei dem die Interessen des Vertrauenden in die Interessen des Vertrauten eingeschlossen sind. Darüber hinaus hat der Vertraute ein Interesse am Fortbestand der Beziehung und damit einen Anreiz sich vertrauenswürdig zu verhalten. Vertrauen ist in diesem Sinne ein rationales, kontextabhängiges Konzept, das spezifische Erwartungen an spezifische Andere voraussetzt und keinesfalls auf Basis generalisierter Erwartungen vergeben wird. HARDIN unterscheidet verschiedene Vertrauenskonstellationen. Zunächst unterscheidet er *ongoing dyadic relationships* von *ongoing group or societal relationships* und differenziert erstere nochmals in *oneway trust relationships* und *mutual trust relationships* aus. *Ongoing dyadic relationships* beziehen sich auf die Mikroebene und beschreiben Vertrauen zwischen einem Vertrauenden und einem Vertrauten, wobei Vertrauen hier sowohl einseitig als auch wechselseitig vergeben werden kann. *Ongoing group or societal relationships* beschreiben Vertrauensbeziehungen auf der Makroebene und stellen in der Regel ein sich überlappendes Beziehungsgeflecht innerhalb von Gemeinschaften dar. (Hardin (2001b), S. 295; Hardin (2002), S. 3ff, S. 9f, S. 13ff, 14ff, S. 21; Hardin (2004), S. 6; Hardin (2004), S. 6f; Hardin (2006), S. 17, S. 19, S. 31)

3.1.2 Definition Sozialen Vertrauens

Ausgehend von den bisherigen Ausführungen lässt sich nun folgende Definition für Soziales Vertrauen finden:

Soziales Vertrauen stellt eine Verhaltensstrategie zum Umgang mit doppelter Kontingenz bzw. zur Überbrückung sozialer Unsicherheit infolge eines Informationsdefizites in Interaktionsbeziehungen dar. Es ist dabei im Wesentlichen auf die Interaktion und Kooperation von Individuen bezogen und ermöglicht eine Reduktion der sozialen Komplexität, die sich aus dem Aufeinandertreffen der Handlungsfreiheit der interagierenden Individuen ergibt. Die

Vertrauensentscheidung nimmt hierzu die unbekannte Zukunft vorweg, indem sie die Erwartungen des Vertrauenden an ein wohlwollendes Handeln des Vertrauten als wahrscheinlichstes Ergebnis der Vertrauensbeziehung annimmt.

Im weiteren Verlauf wird gezeigt werden, dass Soziales Vertrauen mit *trust* und *confidence* zwei voneinander zu unterscheidende Qualitäten annehmen kann, die sich innerhalb der realen Welt als Partikulares und Generalisiertes Vertrauen manifestieren. Dabei wird Partikulares Vertrauen auf der *trust*-Ebene und Generalisiertes Vertrauen auf der *confidence*-Ebene verortet werden.¹⁰ Das resultierende Vertrauensmodell kann dabei als eine integrative Perspektive auf das Phänomen „Vertrauen“ begriffen werden, da es sich aus den Gemeinsamkeiten der im Rahmen dieser Studie vorgestellten Vertrauenskonzeptionen ableitet.

3.2 Generalisiertes Vertrauen

Generalisiertes Vertrauen stellt ebenso wie Partikulares Vertrauen eine Subkategorie Sozialen Vertrauens dar. Diese beiden Vertrauensformen bilden sich diametral gegenüberstehende Endpunkte eines Kontinuums, sind aber mittels bestimmter Mechanismen prinzipiell wechselseitig ineinander überführbar. Der Fokus der nachfolgenden Ausführungen liegt auf theoretischen Ansätzen, die sich mit der Genese Generalisierten Vertrauens beschäftigen. Dabei liegt es in der Natur der Sache, dass eine Definition und Diskussion Generalisierten Vertrauens nicht ohne eine Definition und Diskussion Partikularen Vertrauens auskommt. Entsprechend wird auch Partikulares Vertrauen, wenn auch in einem vergleichsweise limitierten Ausmaß, Gegenstand des nachfolgenden Kapitels sein. Den Schlusspunkt bildet dabei eine Definition beider Vertrauenskategorien.

3.2.1 Determinanten generalisierten Vertrauens

Wie bereits schon die Strukturierung sozialen Vertrauens stellt auch die Systematisierung der Ansätze zur Erklärung der Genese *generalisierten Vertrauens* eine Synthese verschiedener Gliederungsansätze dar. Ziel der nachfolgenden Erörterung ist die Entwicklung einer Gliederung der Determinanten *generalisierten Vertrauens*, die sich an den Gemeinsamkeiten der nachfolgenden Konzeptionen orientiert.

Nach KUNZ „bezieht sich generalisiertes Vertrauen auf die grundsätzliche Vertrauensbereitschaft eines Akteurs, unabhängig von den konkreten Ausprägungen einer bestimmten Situation“ (Kunz (2004), S. 204) und entspricht in diesem Sinne „einer generalisierten Erwartung über die Verlässlichkeit, Wahrhaftigkeit, Fairness und Solidarität der Menschen im Allgemeinen“ (Kunz (2004), S. 204).

Die Bestimmungsfaktoren dieser Form des Vertrauens gliedert KUNZ dabei in *persönlichkeitsorientierte Ansätze* und *strukturorientierte Ansätze*. *Persönlichkeitsorientierte*

¹⁰ Wird im weiteren Verlauf dieser Studie von „Partikularem Vertrauen“ oder „Generalisiertem Vertrauen“ gesprochen, so beziehen sich die Ausführungen auf die im Abschnitt „3.2.5 Generalisiertes Vertrauen vs. Partikulares Vertrauen“ formulierten Definitionen. Andere Schreibweisen (z.B. partikulares Vertrauen, particularized trust bzw. generalisiertes Vertrauen, generalized trust etc.) sind anderen Vertrauenskonzeptionen zu zuschreiben.

3. Theorie: Vertrauen

Ansätze nehmen an, dass insbesondere personelle Faktoren wie persönliche Ressourcen, die individuelle Zufriedenheit und das persönliche Wohlbefinden, individuelle Moralvorstellungen und religiöse Überzeugungen sowie negative und belastende Lebensereignisse Einfluss auf die Entwicklung *generalisierten Vertrauens* nehmen. Die Ausstattung mit persönlichen Ressourcen umfasst dabei in erster Linie die Verfügungsgewalt über sozioökonomisches Kapital, wie Bildung und Einkommen. Zufriedenheit und Wohlbefinden zielt auf eine optimistische Grundeinstellung und ein positives Verhältnis zu anderen Menschen. Moralvorstellungen und religiöse Überzeugungen betreffen die Unterstützung allgemeingültiger Werte und Normen und eine prosoziale Orientierung, während sich der Einfluss negativer Lebensereignisse auf die Neigung zu Argwohn gegenüber den Motiven anderer bezieht. (Kunz (2004), S. 210ff)

Strukturorientierte Ansätze gehen indes von einer zentralen Bedeutung des sozialen und gesellschaftlichen Umfeldes für die Genese *generalisierten Vertrauens* aus. Relevant sind dabei die Einbettung in soziale Netzwerke, die Zufriedenheit mit Demokratie und wirtschaftlicher Lage, die Wahrnehmung, Angehöriger einer Minderheit zu sein, und die Perzeption öffentlicher Sicherheit. Soziale Netzwerke, die auf die Verwirklichung gemeinschaftsorientierter Werte und Normen ausgerichtet sind, tragen dabei zur Internalisierung dieser Werte und Normen bei, die in der Folge auf andere Beziehungen verallgemeinert werden. Eine stabile Demokratie und wirtschaftlicher Wohlstand implizieren ein funktionierendes Rechtssystem und effektive Institutionen, die eine wirksame Bekämpfung opportunistischen Verhaltens ermöglichen und die Wahrnehmung persönlicher Sicherheit fördern. Angehöriger einer Minderheit zu sein und damit verbundene mögliche Diskriminierung sowie die Wahrnehmung mangelnder Sicherheit im sozialen Umfeld kann dagegen zu Zweifeln an der Aufrichtigkeit, Glaubwürdigkeit, Hilfsbereitschaft, Fairness und Vertrauenswürdigkeit anderer Menschen führen. (Kunz (2004), S. 212ff)

ROTHSTEIN/STOLLE verstehen *generalisiertes Vertrauen* als eine Vertrauenseinstellung, die über persönlichen Kontakt hinausgeht und auch solche Personen miteinschließt, die einem nicht unmittelbar bekannt sind. Die Determinanten *generalisierten Vertrauens* lassen sich ihrer Ansicht nach in zwei große Gruppen unterteilen: *society-centered approaches* und *institution-centered accounts*. (Rothstein/Stolle (2002), S. 2; Rothstein/Stolle (2007), S. 2; Rothstein/Stolle (2008b), S. 441)

Society-centered approaches verstehen *generalisiertes Vertrauen* als Folge sozialer Interaktion innerhalb sozialer Netzwerke sowie des kulturellen Erbes. So werden die Mitglieder sozialer Netzwerke in bürgerlichen Werten sozialisiert, die mittels Interaktion mit Nicht-Mitgliedern externalisiert werden. Auf der anderen Seite ergibt sich *generalisiertes Vertrauen* allerdings auch auf Basis historisch-gewachsener bürgerlicher Traditionen. (Rothstein/Stolle (2002), S. 4f; Rothstein/Stolle (2007), S. 4f; Rothstein/Stolle (2008a), S. 275; Rothstein/Stolle (2008b), S. 442)

Institution-centered accounts gehen prinzipiell davon aus, dass die Genese *generalisierten Vertrauens* an den politischen Kontext und formale politische Institutionen, die Administration und die Judikative gebunden ist. Es wird somit angenommen, dass „political institutions create, channel, and influence the amount and type of social capital in their respective societies more than the other way around“ (Rothstein/Stolle (2008a), S. 279). *Institution-centered accounts* lassen sich dabei nochmals in den sogenannten *attitudinal approach* und den sogenannten *institutional-structural approach* differenzieren. Der *attitudinal approach* nimmt eine Korrelation zwischen *generalisiertem Vertrauen* und politischem Vertrauen bzw. Institutionenvertrauen an, wobei die kausale Richtung dieser Verbindung weitestgehend

3. Theorie: Vertrauen

unbestimmt bleibt. Dagegen spielt innerhalb des *institutional-structural approach* insbesondere die Performanz politischer und öffentlicher Institutionen eine große Rolle für die Entwicklung *generalisierten Vertrauens*. Im Fokus stehend sogenannte Implementationsinstitutionen wie Administration und Judikative sowie deren Orientierung an Verhaltensnormen wie Neutralität, Objektivität, Unparteilichkeit oder Fairness. Institutionen fungieren als Spiegelbild der Gesellschaft und können *generalisiertes Vertrauen* generieren, indem sie aufzeigen und durchsetzen, nach welchen Regeln und Prinzipien die jeweilige Gesellschaft funktionieren soll. (Rothstein/Stolle (2002), S. 7ff; Rothstein/Stolle (2007), S. 6ff; Rothstein/Stolle (2008a), S. 278ff; Rothstein/Stolle (2008b), S. 443ff)

PARK beschreibt *generalisiertes Vertrauen* als ein Vertrauen, das sich auf Personen bezieht, die einem persönlich nicht bekannt sind und hinsichtlich deren Vertrauenswürdigkeit einem nur wenige oder keine Informationen vorliegen. Es handelt sich um eine Form des Vertrauens, die nicht unmittelbar auf der Vertrauenswürdigkeit spezifischer Menschen basiert, wenn auch nicht ausgeschlossen ist, dass vergangene Erfahrungen eine direkte oder indirekte Rolle spielen können. (Park (2012), S. 3)

PARK unterscheidet insgesamt drei Ansätze, die die Entstehung *generalisierten Vertrauens* zu erklären versuchen. *Network-based accounts* stellen die Relevanz sozialer Netzwerke für die Genese *generalisierten Vertrauens* hervor. Innerhalb solcher Netzwerke werden Kooperationsnormen und Vertrauen kultiviert, die in der Folge auf die gesamte Gesellschaft übergreifen. (Park (2012), S. 4)

Institutions-based accounts sehen die Quelle *generalisierten Vertrauens* im Handeln und der Leistungsfähigkeit staatlicher Institutionen. Dabei besteht allerdings Uneinigkeit, ob staatliche Intervention per se einen negativen Effekt auf *generalisiertes Vertrauen* hat oder ob nicht insbesondere Institutionen zur Bewahrung der öffentlichen Ordnung oder wohlfahrtsstaatliche Institutionen förderlich auf die Entwicklung *generalisierten Vertrauens* einwirken. (Park (2012), S. 5f)

Culture-based accounts machen spezifische Werteinstellungen und Weltanschauungen für die Entwicklung *generalisierten Vertrauens* verantwortlich. So ist *generalisiertes Vertrauen* eng verbunden mit einem optimistischen und egalitären Weltbild, das im Rahmen der Sozialisation ausgebildet bzw. durch die Eltern vererbt wird. (Park (2012), S. 6f)

Wie bereits beschrieben betrachtet NANNSTAD *partikulares* und *generalisiertes Vertrauen* als einander gegenüberstehende Extrempunkte eines Vertrauenskontinuums auf einer zweiten Differenzierungsebene sozialen Vertrauens. Während *partikulares Vertrauen* als Vertrauen gegenüber einer bestimmten Person, über die wir Informationen besitzen, bezüglich einer bestimmten Angelegenheit definiert ist, beschreibt *generalisiertes Vertrauen* ein Vertrauen, das auch auf Fremde ausgeweitet ist. (Nannestad (2008), S. 414f)

NANNSTAD unterscheidet im Wesentlichen vier Determinanten *generalisierten Vertrauens*. Die *civic society explanation* geht davon aus, dass die bürgerliche Gesellschaft über die Vermittlung bürgerlicher Tugenden innerhalb sozialer Netzwerke Einfluss auf die Genese *generalisierten Vertrauens* ausübt. Weitergegeben werden dabei neben Vertrauensnormen auch Reziprozitäts- und Kooperationsnormen. (Nannestad (2008), S. 422f; Park (2012), S. 4)

Die *institutional explanation* fokussiert auf die Qualität politischer Institutionen und deren Vermögen, Anreize zu schaffen, sich vertrauenswürdig zu verhalten. Institutionen sind in der Lage, Sanktionen gegen Personen auszusprechen und durchzusetzen, die einen

3. Theorie: Vertrauen

Vertrauensmissbrauch begangen haben, und verringern auf diese Weise das Risiko der Vertrauensvergabe bzw. fungieren als Substitute für Vertrauen. Gleichzeitig tragen sie durch institutionelle Fairness und Gleichbehandlung aller Bürger zur Etablierung von Vertrauensnormen bei. (Nannestad (2008), S. 423ff; Park (2012), S. 4)

Die *cultural-values explanation* stellt per Sozialisation vermittelte kulturelle Normen, eine optimistische Weltanschauung sowie religiöse und egalitäre Werte als Bedingungsfaktoren *generalisierten Vertrauens* in den Vordergrund. (Nannestad (2008), S. 425f; Park (2012), S. 4)

Die *social-structural explanation* macht ethnische Diversität bzw. Heterogenität als Determinante *generalisierten Vertrauens* aus. Dabei bleibt zunächst unklar, ob Diversität einen positiven oder negativen Effekt auf Vertrauen hat. Insofern als die ethnische Diversität die soziale Distanz zwischen den Bürgern einer Gesellschaft vergrößert, wirkt Diversität eher nachteilig auf *generalisiertes Vertrauen*. Auf der anderen Seite kann die Durchmischung verschiedener Ethnien auch dazu führen, dass Vorurteile abgebaut werden und interethnisches Vertrauen entsteht. (Nannestad (2008), S. 426ff; Park (2012), S. 4)

Betrachtet man die hier vorgestellten Systematisierungen der Determinanten *generalisierten Vertrauens*, so lassen sich neben Parallelen in den jeweiligen Konzeptionen auch unterschiedliche Differenzierungen der Bedingungsfaktoren *generalisierten Vertrauens* beobachten. Gemeinsam ist dabei allen Ansätzen die Betonung der Relevanz von bürgerlicher Gesellschaft, politischen Institutionen, kulturellem Erbe bzw. Sozialisation, spezifischen Wertüberzeugungen und einer optimistischen Weltanschauung für die Genese *generalisierten Vertrauens*. KUNZ und NANNESTAD ergänzen diese Aspekte zusätzlich noch durch den Einfluss ethnischer Diversität.

Während KUNZ die genannten Bedingungsfaktoren *generalisierten Vertrauens* in ein verhältnismäßig sparsames Gliederungsmodell einordnet, finden sich bei den übrigen Theoretikern differenziertere Systematisierungsvorschläge. Besonders ähnlich sind sich dabei die Konzeptionen von PARK und NANNESTAD, wobei zu beachten ist, dass die Systematisierung PARKs aus der von NANNESTAD hervorgeht und sich im Wesentlichen nur durch die Aussparung ethnischer Diversität von dieser unterscheidet. NANNESTADs Konzeption stellt hiernach die vergleichsweise ausdifferenzierteste Systematisierung der Determinanten *generalisierten Vertrauens* dar und soll als Grundlage für die hier zu entwickelnde Gliederung dienen. Ein Vergleich von NANNESTAD und ROTHSTEIN/STOLLE bringt zutage, dass der Gliederungspunkt *society-centered approaches* im Großen und Ganzen der *civic society explanation* und der *cultural-values explanation* bei NANNESTAD entspricht, weshalb letztere als Ausdifferenzierung dieser übergeordneten Kategorie betrachtet werden. *Institutional explanation* bei NANNESTAD und *institution-centered accounts* bei ROTHSTEIN/STOLLE scheinen sich ebenfalls zu entsprechen, wobei in diesem Fall ROTHSTEIN/STOLLE mit dem *attitudinal approach* und dem *institutional-structural approach* eine zweite Differenzierungsebene einfügen. Die *social-structural explanation* bzw. ethnische Diversität soll einstweilen als eigenständige Kategorie auf der gleichen Ebene wie die *society-centered approaches* und die *institution-centered accounts* behandelt werden.

Auf diese Weise entsteht eine zwei Ebenen umfassende Systematisierung der Determinanten *generalisierten Vertrauens*. Auf der ersten Ebene und parallel zueinander befinden sich die *society-centered approaches*, die *institution-centered accounts* und die *social-structural explanation* bzw. ethnische Diversität. Die *society-centered approaches* differenzieren sich auf einer untergeordneten Ebene in *civic society explanation* und *cultural-values explanation* auf,

3. Theorie: Vertrauen

während sich *institution-centered accounts* in den *attitudinal approach* und den *institutional-structural approach* aufspalten. Im Kontext der *social-structural explanation* bzw. ethnischer Diversität konkurrieren mit der Kontakt-, Konflikt- und Konstrukttheorie drei Erklärungsansätze.

Tab. 3.1: Systematisierung der Determinanten *generalisierten Vertrauens*

<u>Society-centered accounts</u> Generalisiertes Vertrauen ergibt sich aus den historischen und kulturellen Erfahrungen einer Gesellschaft, wobei formelle und informelle Vereinigungen als Katalysator dienen.	<u>Institution-centered accounts</u> Generalisiertes Vertrauen ist eingebettet in und verbunden mit politischen und judikativen Institutionen.	<u>Ethnische Heterogenität</u> Abgrenzungstendenzen zwischen ethnischen Gruppen schaffen soziale Distanz, was gleichermaßen Einfluss auf intraethnisches und interethnisches Vertrauen (<i>generalisiertes Vertrauen</i>) ausübt.
Networks-based approaches Generierung von <i>generalisiertem Vertrauen</i> als Folge von <i>face-to-face</i> -Interaktion in Interessenverbänden bei denen eine Sozialisation anhand von Vertrauens- und Reziprozitätsnormen stattfindet. Culture-based approaches <i>Generalisiertes Vertrauen</i> ist Folge der Sozialisation in einem Umfeld, das eine optimistische Weltanschauung propagiert.	Attitudinal-approaches Institutionelles Vertrauen und <i>generalisiertes Vertrauen</i> wirken aufeinander bzw. hängen zusammen. Institutional-structural approaches Institutionen tragen zur Generierung <i>generalisierten Vertrauens</i> bei, indem sie anomisches Verhalten sanktionieren sowie über die Verbreitung von kulturellen Prinzipien und Normen die Werte und Überzeugungen der Menschen gestalten.	Kontakttheorie Ethnische Vielfalt trägt dazu bei, Vorurteile abzubauen sowie wechselseitige Toleranz und Solidarität zu fördern, was eine Abschwächung von <i>ingroup-outgroup</i> -Unterschieden bewirkt und <i>outgroup</i> -Vertrauen fördert. Konflikttheorie Aus ethnischer Diversität ergeben sich Auseinandersetzungen um knappe sozioökonomische Ressourcen, was die Feindschaft gegenüber der <i>outgroup</i> fördert und negative Konsequenzen für das <i>generalisierte Vertrauen</i> hat. Konstrukttheorie Ethnische Heterogenität übt sowohl auf das <i>ingroup</i> - als auch auf das <i>outgroup</i> -Vertrauen einen negativen Einfluss aus, was in einem Rückzug aus dem sozialen Leben resultiert („ <i>hunkering down</i> “).
Kritik^a <ul style="list-style-type: none"> • <i>ingroup</i>-Vertrauen vs. <i>outgroup</i>-Misstrauen • Mechanismus der Übertragung auf gesamte Gesellschaft • Kausale Richtung der Genese • Historischer Determinismus 	Kritik^a <ul style="list-style-type: none"> • Einstellungen gegenüber Institutionen sind nicht mit deren eigentlichen Charakteristika verbunden. • Versch. Varianten des Institutionenvertrauens werden unter ein Label subsumiert. • Kausaler Mechanismus ist unklar. • Repräsentations- vs. Implementationsinstitutionen. 	Kritik <ul style="list-style-type: none"> • Effekt ethnischer Heterogenität bleibt in letzter Instanz unbestimmt.

^a Vgl. Abschnitt „3.2.3.2.1 Vertrauensnormen als Folge von „street-level-bureaucracy“ (Rothstein/Stolle)“

(Quelle: Gundelach (2014a); Gundelach (2014b); Newton (1999); Putnam (2007); Rothstein/Stolle (2002); Rothstein/Stolle (2007))

3.2.2 Society-centered Accounts

Gesellschaftszentrierte Ansätze gehen davon aus, dass sich *generalisiertes Vertrauen* aus den historischen und kulturellen Erfahrungen einer Gesellschaft ergibt, wobei formelle und informelle Vereinigungen als Katalysator fungieren. *Generalisiertes Vertrauen* stellt dann die Folge Sozialen Kapitals dar, welches sich über einen längeren Zeitraum, teilweise über Jahrhunderte, aufgebaut hat, und ist somit stark im jeweiligen historischen und kulturellen Kontext verwurzelt. Motor der Genese *generalisierten Vertrauens* ist hierbei die regelmäßige soziale Interaktion der Gesellschaftsmitglieder.

Elementare Kritik an den gesellschaftszentrierten Ansätzen ergibt sich aus dem propagierten historischen Determinismus der Genese *generalisierten Vertrauens*. Die Entstehung *generalisierten Vertrauens* erfolgt in dieser Lesart über lange Zeiträume, die empirisch nur schwer zu erfassen bzw. im Detail zu prüfen sind. Gleichzeitig ist schwer nachzuvollziehen, wie eine langfristig gewachsene und entsprechend stabile Einstellung wie *generalisiertes Vertrauen*, die zudem in die sozialen Strukturen der jeweiligen Gesellschaft eingebettet ist, in der Gegenwart durch einzelne Ereignisse zersetzt werden kann.

3.2.2.1 Network-based Approach

Netzwerk-basierte Ansätze erklären die Genese *generalisierten Vertrauens* als Folge der Mitgliedschaft in sozialen Netzwerken und der damit verbundenen Sozialisationsprozesse. So vermitteln soziale Netzwerke bürgerliche Tugenden wie Vertrauen, Reziprozität und Kooperationsbereitschaft sowie demokratische Werte und Normen, indem sie die Interaktion zwischen Menschen fördern und als Umgebung zum Knüpfen und zur Pflege von Kontakten fungieren. Die positiven Interaktionserfahrungen mit anderen Netzwerkmitgliedern werden dann auf Nicht-Mitglieder übertragen. Es wird dabei angenommen, dass positive Erlebnisse mit Vertrauen innerhalb sozialer Netzwerke und die sich daraus ergebenden Vorzüge, Netzwerkmitglieder gegenüber Fremden nach dem Prinzip „*in dubio pro reo*“ handeln lässt, bis sich diese als nicht vertrauenswürdig herausgestellt haben. Es wird außerdem erwartet, dass Personen, die in mehrere soziale Netzwerke eingebunden sind, eher bereit sind Vertrauen zu schenken, als solche, die nicht in sozialen Netzwerken aktiv sind, da erstere mehr Erfahrungen mit dem Austausch von Gefälligkeiten und Reziprozität gemacht haben. (Nannestad (2008), S. 422f; Park (2012), S. 4)

Unklar bleibt allerdings, nach welchem Mechanismus die Erweiterung des *partikularen Vertrauens* zwischen Netzwerkmitgliedern auf Nicht-Mitglieder in verschiedenen Kontexten funktioniert. So ist nicht geklärt, welche Aspekte sozialer Interaktion in sozialen Netzwerken mit der Genese *generalisierten Vertrauens* verbunden sind und es existiert bisher kein Erklärungsmodell dafür, wie *generalisiertes Vertrauen* durch soziale Interaktion auf der Mikroebene geschaffen wird. (Nannestad (2008), S. 422f; Park (2012), S. 4; Rothstein/Stolle (2008a), S. 277f)

Des Weiteren wird der Sozialisationseinfluss von sozialen Netzwerken angezweifelt. Hierbei wird argumentiert, dass die in Netzwerken verbrachte Zeit verglichen mit anderen Aktivitäten in anderen sozialen Kontexten, vergleichsweise gering ausfällt und sie somit kaum prägenden Einfluss auf das Verhalten der Mitglieder haben können. Außerdem scheint im Rahmen des Beitritts zu sozialen Netzwerken eine Selbstselektion stattzufinden, wonach sogenannte „*truster*“ eher Netzwerkmitglieder werden als sogenannte „*distruster*“. Mitglieder in sozialen

3. Theorie: Vertrauen

Netzwerken weisen somit bereits im Vorhinein ihres Beitritts einen eher hohen Vertrauenslevel auf, was die kausale Richtung der Genese *generalisierten Vertrauens* aus sozialen Netzwerken heraus in Frage stellt. Vielmehr scheint eher *generalisiertes Vertrauen* das *partikulare Vertrauen* zwischen den einzelnen Netzwerkmitgliedern zu bewirken. (Rothstein/Stolle (2002), S. 5; Rothstein/Stolle (2007), S. 4f; Rothstein/Stolle (2008a), S. 277f; Rothstein/Stolle (2008b), S. 442)

Ebenso muss die programmatische Ausrichtung der sozialen Netzwerke beachtet werden. So kann hinterfragt werden, ob die Partizipation innerhalb sozialer Netzwerke tatsächlich auf die Entwicklung *generalisierten Vertrauens* und die Bereitschaft zur Kooperation mit Nicht-Mitgliedern gerichtet ist oder ob nicht vielmehr eine Exklusion von Nicht-Mitgliedern im Sinne einer Abgrenzung von der *outgroup* stattfindet. So scheinen soziale Netzwerke eher dafür zu sorgen, dass die Mitglieder ihren Kameraden mehr Vertrauen entgegenbringen und lediglich die Kooperationsbereitschaft hinsichtlich der Ziele der eigenen Gruppe gestärkt wird, während gegenüber Nicht-Mitgliedern Misstrauen etabliert wird. Unterschieden werden muss in diesem Zusammenhang das sogenannte *bonding* (Kontakt mit Menschen wie man selbst) und *bridging* (Kontakt mit vielen unterschiedlichen Menschen). Vereinigungen, die eher auf *bridging* ausgerichtet sind, ermöglichen auch Kontakte über die Gruppengrenzen des eigenen Netzwerkes hinaus und wären deshalb eher dazu in der Lage, einen Beitrag zur Entwicklung *generalisierten Vertrauens* zu leisten. Wohingegen sich auf *bonding* ausgerichtete Vereinigungen auf Kontakte innerhalb der eigenen Gruppengrenzen beschränken. Eine andere Trennlinie könnte zwischen solchen Netzwerken gezogen werden, die das Ziel einer integrativen Gesellschaft und die Bewahrung von Kollektivgütern bezwecken, und solchen Netzwerken, die eher an Status, Gruppenidentität und materiellen Gütern orientiert sind. In beiden Fällen würden letztere nur wenig zur Entstehung *generalisierten Vertrauens* beitragen. (Rothstein/Stolle (2002), S. 5f; Rothstein/Stolle (2007), S. 5f; Rothstein/Stolle (2008a), S. 276; Rothstein/Stolle (2008b), S. 442f)

3.2.2.1.1 Vertrauen und Soziales Kapital (Putnam)

Ausgangspunkt von PUTNAMs Beschäftigung mit „Vertrauen“ ist eine Metaanalyse der Verteilung *sozialen Kapitals* in modernen Gesellschaften. Dabei stellt „Vertrauen“ seiner Auffassung „an essential component of social capital“ (Putnam/Leonardi/Nannetti (1993), S. 170) dar. Sein Ansatz ist deshalb nicht unbedingt als eine eigenständige Vertrauentheorie zu verstehen, sondern setzt ein bereits bestehendes theoretisches Verständnis von „Vertrauen“ voraus. „Vertrauen“ bleibt in den Erklärungsrahmen des Begriffes des *Sozialen Kapitals* eingebettet und wird quasi als interindividueller Kitt beschrieben, der eine Gesellschaft als Gemeinschaft zusammenhält:

„Social capital here refers to features of social organization, such as trust, norms, and networks, that can improve the efficiency of society by facilitating coordinated actions.“ (Putnam/Leonardi/Nannetti (1993), S. 167)

Nach PUTNAMs Ansicht befinden sich soziale Netzwerke, soziales Vertrauen sowie soziale Werte und Normen in einem ständigen wechselseitigen Verstärkungszirkel und tragen dabei zur Etablierung und Zementierung von „*civic virtues*“ bei. (Putnam (2000), S. 19, S. 136f) Soziale Netzwerke, welche sich aus zivilgesellschaftlichem Engagement ergeben, ermöglichen Interaktion und Kommunikation zwischen den einzelnen Individuen, sodass Informationen übereinander gesammelt werden können. Dies wirkt sich positiv auf die wechselseitige

3. Theorie: Vertrauen

Zuschreibung von Vertrauenswürdigkeit aus, was wiederum großen Einfluss auf die innergesellschaftliche Kooperationsbereitschaft hat:

„In communities where people can be confident that trusting will be requited, not exploited, exchange is more likely to ensue.“ (Putnam/Leonardi/Nannetti (1993), S. 172)

Auch ein Rechtssystem und seine exekutiven Institutionen können als Moderatoren von Vertrauenswürdigkeit dienen, sind aber keineswegs so effektiv wie soziales Vertrauen, das sich aus sozialen Netzwerken generiert. Während soziales Vertrauen auf Basis eines Rechtssystems zu exorbitanten Transaktionskosten führt, schafft es soziales Vertrauen aus sozialen Netzwerken, diese wirksam zu reduzieren. (Putnam (2000), S. 135f)

Bei sozialem Vertrauen aus sozialen Netzwerken ist zu unterscheiden, ob Vertrauenswürdigkeit aus persönlicher Erfahrung oder aus einer allgemeinen Gesellschaftsnorm resultiert. Ersteres wird dabei durch den Begriff *thick trust* beschrieben und meint „Trust embedded in personal relations that are strong, frequent, and nested in wider networks (...)“ (Putnam (2000), S. 136). Dagegen beschreibt *thin trust* ein Vertrauen in den *generalisierten Anderen*. PUTNAM stellt hierbei fest, dass Vertrauen in einen *generalisierten Anderen* den Vertrauensradius einer Person erweitert, was mit einer Vergrößerung des Handlungsspielraumes dieser Person einhergeht. Entsprechend stellt auch *thin trust* das eigentliche soziale Vertrauen dar, wie PUTNAM anhand der Vertrauensdefinition von RAHN/TRENSUE aufzeigt (Putnam (2000), S. 136):

„Social, or generalized, trust can be viewed as a 'standing decision' to give most people – even those whom one does not know from direct experience – the benefit of the doubt.“ (Rahn/Trensue (1998), S. 545)

Vertrauen begünstigt Kooperation, die wiederum neues Vertrauen schafft bzw. bestehendes erweitert. Gesellschaften, die soziales Vertrauen verinnerlicht haben, sind deshalb effizienter als solche, in denen Misstrauen vorherrscht. Soziales Vertrauen kondensiert in ersteren in einem übergeordneten „principle of generalized reciprocity“ (vgl. Putnam (2000), S. 134ff). Hierbei ist zu beachten, dass man zwei Formen von Reziprozität unterscheiden muss. *Balanced* bzw. *specific reciprocity* bezieht sich auf einen „simultaneous exchange of items of equivalent value“ (Putnam/Leonardi/Nannetti (1993), S. 172) oder etwas griffiger ausgedrückt: „I'll do this for you if you do that for me“ (Putnam (2000), S. 20). Auf diese Art und Weise verstandene Reziprozität beschreibt somit, dass auf einen erwiesenen Gefallen in unmittelbarer zeitlicher Nähe ein wertkongruenter Gegengefallen erfolgen muss. *Generalized* bzw. *diffuse reciprocity* ist dagegen auf eine „continuing relationship of exchange that is at any given time unrequited or imbalanced, but that involves mutual expectations that a benefit granted now should be repaid in the future“ (Putnam/Leonardi/Nannetti (1993), S. 172) bezogen. An anderer Stelle drückt PUTNAM dies wie folgt aus:

„I'll do this for you now, without expecting anything immediately in return and perhaps without even knowing you, confident that down the road you or someone else will return the favor.“ (Putnam (2000), S. 134)

Generalized reciprocity kennzeichnet demnach eine Einstellung, bei der auf lange Sicht auf einen Ausgleich zwischen erwiesenen Gefallen und Gegengefallen gesetzt wird. Von wem diese Gegengefallen erwiesen werden, ist dabei irrelevant, vielmehr wird sich darauf verlassen, dass die Gleichung am Ende aufgeht. In Anlehnung an TAYLOR handelt es sich bei Reziprozität dann um einzelne altruistische Handlungen, die durch ihre charakteristische Wechselseitigkeit für alle

Teilhaber am auf Reziprozität basierenden System einen Vorteil generiert. (Putnam (2000), S. 134f; Putnam/Leonardi/Nannetti (1993), S. 172; Taylor (1982), S. 28f)

Zivilgesellschaftliches Engagement in sozialen Netzwerken hat demnach ausgesprochen positive Konsequenzen für eine Gesellschaft. Es erhöht den innergesellschaftlichen Kommunikationslevel, was es den Teilhabern der Netzwerke ermöglicht, eine vertrauenswürdige Reputation zu erwerben und sie so wechselseitig interessanter für Kooperationsvorhaben macht. Personen, die einen hohen Grad an Vertrauenswürdigkeit erreicht haben, sind dabei gegenüber Personen mit einer niedrigeren Vertrauenswürdigkeit im Vorteil. Gleichzeitig reduziert die Teilhabe an sozialen Netzwerken die Gefahr von Untreue bzw. Vertrauensmissbrauch. Am Ende steht die Etablierung einer Reziprozitätsnorm, die eine Vertrauensgrundlage darstellt, was wiederum die Bildung sozialer Netzwerke fördert. (Putnam (1995a), S. 67, S. 73; Putnam (1995b), S. 20, S. 29; Putnam/Leonardi/Nannetti (1993), S. 169, S. 173f)

Soziales Vertrauen kann nach PUTNAM somit aus zwei Quellen generiert werden: Aus (1) einer Reziprozitätsnorm und aus (2) Netzwerken zivilgesellschaftlichen Engagements. (Putnam/Leonardi/Nannetti (1993), S. 171)

3.2.2.1.2 Vertrauen als rationale Entscheidung (Herreros)

HERREROS versteht Vertrauen als rationale Entscheidung, die mit einem Risiko verbunden ist, das sich aus der Tatsache ergibt, dass die Vertrauenswürdigkeit des Interaktionspartners nicht sichergestellt ist. Im engeren Sinne beschreibt diese Entscheidung dabei nicht eigentliches Vertrauen, sondern die Bereitschaft jemandem anderen etwas anzuvertrauen, sodass Vertrauen wie folgt definiert wird (Herreros (2004), S. 7f):

„Trust is reflected in the expectation about the other individual's trustworthiness (p). Trust, therefore, is a more or less well-grounded expectation about the preferences of other people.“ (Herreros (2004), S. 8)

Vertrauen ist dabei eine Reaktion auf soziale Unsicherheit, die sich aus einem Informationsdefizit hinsichtlich des Verhaltens und Handelns des Interaktionspartners ergibt. Gleichzeitig ist dieses Informationsdefizit funktional für die Entstehung von Vertrauen, da dieses nur benötigt wird, wenn die Informationsdichte bezüglich einer anderen Person gering ist. Sind genug Informationen zugänglich, so wird kein Vertrauen benötigt, um soziale Unsicherheit zu überwinden. Allerdings benötigt Vertrauen einen kritischen Level an Information über das Gegenüber, um bestimmen zu können, ob ihm vertraut werden kann oder nicht. (Herreros (2004), S. 8)

Innerhalb des Vertrauens unterscheidet HERREROS in *particularized trust* und *social bzw. generalized trust*.

Particularized trust beschreibt ein Vertrauen „between persons that know each other“ (Herreros (2004), S. 21) bzw. „trust in known people“ (Herreros (2004), S. 27). Solches Vertrauen ergibt sich als Folge wiederholter Interaktion und basiert auf in vergangenem Verhalten begründeten Erwartungen hinsichtlich des zukünftigen Verhaltens des Interaktionspartners. (Herreros (2004), S. 28)

Social bzw. generalized trust ist hingegen wie folgt definiert:

3. Theorie: Vertrauen

„Social or generalized trust is trust in unknown people about whom no information about their trustworthiness is available.“ (Herreros (2004), S. 23)

Die Entstehung solchen Vertrauens erklärt Herreros aus der Perspektive des Vertrauten und nimmt an, dass dieser Interesse an einer vertrauenswürdigen Reputation und einem Selbstbild als vertrauenswürdige Person hat. So könnte ein Vertrauensmissbrauch die vertrauenswürdige Reputation beschädigen, wenn man bei diesem Vertrauensmissbrauch von Personen aus dem persönlichen Bekanntenkreis beobachtet wird. Ebenso würde ein Vertrauensmissbrauch das positive Selbstbild beschädigen. (Herreros (2004), S. 45f)

Sowohl *particularized trust* als auch *social* oder *generalized trust* werden nicht explizit herbeigeführt, sondern entstehen als Nebenprodukt anderer Aktivitäten. (Herreros (2004), S. 27, S. 51ff)

3.2.2.1.3 Thick und Thin Trust (Wollebæk/Selle)

WOLLEBÆK/SELLE gehen von der Prämisse aus, dass soziales Vertrauen durch wiederholte *face-to-face*-Interaktion mit Personen in sozialen Vereinigungen generiert wird. Dabei wirkt sich insbesondere der Kontakt mit Personen, die aus unterschiedlichen sozialen Verhältnissen stammen, positiv auf die Entwicklung von Vertrauen aus. Die innerhalb solcher Gemeinschaften erschaffenen Netzwerke ermöglichen darüber hinaus eine Verbreitung des Vertrauens über die gesamte Gesellschaft. (Wollebæk/Selle (2002), S. 32; Wollebæk/Selle (2003), S. 67; Wollebæk/Selle (2007), S. 1)

Soziales Vertrauen kann sich prinzipiell in zwei Formen manifestieren, wobei der Grad der sozialen Abgeschlossenheit bzw. Offenheit eines Netzwerkes darüber entscheidet, welche Form sich entwickeln kann. So entsteht *thick trust* vornehmlich in Vereinigungen, die ein hohes Maß an sozialer Geschlossenheit aufweisen. Solches Vertrauen ist eingebettet in die sozialen Beziehungen innerhalb des sozialen Netzwerkes und beschränkt sich ausschließlich auf dessen unmittelbare Mitglieder. (Wollebæk/Selle (2002), S. 33; Wollebæk/Selle (2007), S. 2f;) *Thin trust* hingegen beschreibt „a *thin*, generalized social trust between individuals who do not know each other very well, if at all“ (Wollebæk/Selle (2002), S. 33). Solches Vertrauen nimmt an, dass die meisten Menschen vertrauenswürdig sind und im Geiste eines gemeinnützigen Interesses handeln werden. Es überbrückt Gruppen- bzw. Netzwerkgrenzen und stellt eine wichtige Ressource bei der Interaktion mit Fremden dar. (Wollebæk/Selle (2007), S. 2f)

Neben diesem als *bonding* (Beschränkung von Vertrauen auf Netzwerkmitglieder) bzw. *bridging* (Erweiterung von Vertrauen auf Personen außerhalb des Netzwerkes) bezeichneten Effekt von sozialen Netzwerken ergeben sich auch sogenannte *externe* und *interne* Effekte. (Wollebæk/Selle (2002), S. 35; Wollebæk/Selle (2003), S. 68; Wollebæk/Selle (2007), S. 2f) *Externe* Effekte bezeichnen die Einflussnahme der Netzwerkmitglieder auf gesellschaftliche und politische Entscheidungsprozesse. *Interne* Effekte beschreiben die Sozialisation der Netzwerkmitglieder durch das Netzwerk. So werden Fähigkeiten und Kompetenzen in den Bereichen Kooperation, Solidarität und Gemeinschaftssinn vermittelt. Für Vertrauensentwicklung scheinen dabei vornehmlich diese *internen* Effekte verantwortlich zu sein.

WOLLEBÆK/SELLE machen nun drei Mechanismen aus, mit deren Hilfe sich Vertrauen aus sozialen Netzwerken auf die Gesellschaft übertragen lässt (Wollebæk/Selle (2002), S. 37f Wollebæk/Selle (2007), S. 4):

- (1) *Cognitive inference*: Erfahrungen mit sozialen Netzwerken und ihrer Funktionsweise beeinflussen unsere Wahrnehmung hinsichtlich der Kooperationsbereitschaft anderer in positiver Art und Weise. Je erfolgreicher und sichtbarer die jeweilige soziale Vereinigung innerhalb der Gesellschaft ist, umso stärker ist dieser Effekt.
- (2) *Intermediary structures*: Soziale Netzwerke verbinden Individuen mit der Gesellschaft. Sie stellen eine organisationale Infrastruktur bereit, die sich sowohl Mitglieder wie auch Nichtmitglieder nutzbar machen können, wenn sie sie benötigen.
- (3) *Moderating and cumulative effect*: Überlappende und ineinandergreifende soziale Netzwerke erhöhen die Wahrscheinlichkeit, dass Menschen unterschiedlicher Couleur aufeinandertreffen. Das Ergebnis ist oft die Etablierung neuer, quer zu den Ursprungsnetzwerken verlaufender Netzwerke, die die Toleranz und das Verständnis füreinander verstärken. Die Interaktion mit Personen mit unterschiedlichsten sozialen und politischen Hintergründen und entsprechenden Zielen und Präferenzen zwingt dazu die eigenen Einstellungen anzupassen, um diese sozialen Beziehungen aufrechtzuerhalten.

Daneben bedeutet die Mitgliedschaft in multiplen Netzwerken mehr und breiter gestreute Interaktion, was sich ebenfalls positiv auf die Entwicklung von Vertrauen auswirkt.

Letztendlich kommen WOLLEBÆK/SELLE auf der Basis multipler empirischer Untersuchungen allerdings zu dem Schluss, dass die Sozialisation innerhalb sozialer Netzwerke nur in geringem Maße zur Generierung sozialen bzw. *generalisierten Vertrauens* beiträgt (Wollebæk/Selle (2007), S. 21) bzw. dass „(...) the role of associations as generators or catalysts of trust, or networks of civic engagement, is subordinated to their role as institutions expressing and sustaining common values and resources“ (Wollebæk/Selle (2004), S. 251). Soziale Netzwerke sollten vielmehr als „institutionally-embedded stores of trust, norms and networks of civic engagement, rather than as generators, catalysts or vehicles“ (Wollebæk/Selle (2004), S. 252) angesehen werden. Soziale Netzwerke sorgen in diesem Sinne mehr für eine Institutionalisierung sozialen bzw. *generalisierten Vertrauens* als für dessen Generierung. (Wollebæk/Strømsnes (2008), S. 260)

3.2.2.1.4 Vertrauen aus sozialen Netzwerken (Stolle)

Während STOLLE zunächst den Fokus auf die Rolle von sozialen Vereinigungen bei der Generierung *generalisierten Vertrauens* legt, gewinnen in späteren Beiträgen auch die Bedeutung der familiären Sozialisation sowie die Beeinflussung durch staatliche Institutionen an Relevanz. STOLLE entwickelt in dieser Folge eine integrative Perspektive auf die verschiedenen Faktoren der Entwicklung *generalisierten Vertrauens*.

STOLLE unterscheidet zunächst zwei Arten von Vertrauen, *generalized interpersonal trust* und *personal, private* bzw. *particularized trust*, die wie folgt spezifiziert sind:

„(...) we must examine the distinction between generalized trust, which extends beyond the boundaries of face-to-face interaction, and private or personalized trust, which involves people personally known. Personalized trust results from cooperation experiences and repeated interaction with the immediate circle of known people, whether that be family, community, or fellow members of a voluntary association, and is also directed to them. Generalized interpersonal trust, on the other hand is a trust that goes beyond the boundaries of kinship and friendship, and even beyond the boundaries of acquaintance. It can act as a social lubricant that makes a variety of forms of social interaction and cooperation possible in a wider community or region (...).“ (Stolle (1998), S. 503)

Außerdem heißt es an weiteren Stellen:

3. Theorie: Vertrauen

„Generalized attitudes of trust extend beyond the boundaries of face-to face interaction and incorporate people not personally known. They are indicated by an abstract preparedness to trust others and to engage in actions with others. These attitudes of trust are generalized when they go beyond specific personal settings in which the partner to be cooperated with is already known; they even go beyond the boundaries of kinship, friendship, and acquaintance. In this sense the scope of generalized trust should be distinguished from trust in personal relationships. This more immediate form of trust may be called private or personalized trust; it results from cooperation and repeated interactions with one's immediate circle, whether that be a family, community, or voluntary association.“ (Stolle (2001a), S. 205)

„Also trust has various forms: particularized trust, for example, is directed at members of one's actual or imagined group and it provides the necessary fabric for achieving group goals more efficiently. Whereas generalized trust is a rather encompassing value that reaches to include most people and is most beneficial in the contact with strangers.“ (Stolle/Hooghe (2003), S. 28)

Nach STOLLE ermöglichen soziale Vereinigungen die Entstehung und Entwicklung *generalisierten Vertrauens*, indem sie unter ihren Mitgliedern kooperative Einstellungen und kooperatives Verhalten fördern. Hierbei spielt insbesondere die vermehrte *face-to-face*-Interaktion zwischen den Vereinigungsmitgliedern eine wesentliche Rolle. Sie ermöglicht es den Mitgliedern, wechselseitig Erfahrungen mit und Wissen über andere Gruppenmitglieder zu erlangen, was zu einer größeren Bereitschaft zu kollektivem Handeln und Kooperation sowie der Entwicklung von *ingroup*-Vertrauen führt. Durch das Vertrauen innerhalb der Vereinigung wird dann die Verwirklichung von gemeinsamen Zielen vereinfacht und effizienter gestaltet. (Stolle (1998), S. 500; Stolle (2001a), S. 204f; Stolle (2001b), S. 119; Stolle (2003), S. 24f; Stolle/Rochon (1999), S. 192f)

In der Folge werden diese positiven Gruppenerfahrungen auf Gesellschaft als Ganzes übertragen und sorgen dort für die Entwicklung von „cooperation between all people for all sorts of purposes“ (Stolle (2003), S. 205). Für den Mechanismus, der bei der Generalisierung des *ingroup*-Vertrauens wirksam wird, existieren nach STOLLE zwei Erklärungsmodelle. Zum einen könnte erfahrungsbasiertes Vertrauen aus der regelmäßigen Interaktion mit bekannten Personen ab einem bestimmten Punkt auf Personen übertragen werden, welche einem nicht persönlich bekannt sind. *Partikulares Vertrauen* und *generalisiertes Vertrauen* würden hiernach positiv aufeinander wirken. (Stolle (1998), S. 503)

Gegen diese These spricht, dass die Entstehung sogenannter *strong ties* innerhalb von sozialen Netzwerken eher hemmend auf die Entwicklung von *generalisiertem Vertrauen* wirken. So verhindern enge Bindungen und Verpflichtungsbeziehungen in der Regel Offenheit gegenüber der äußeren Umwelt. Entsprechend geht das zweite Erklärungsmodell davon aus, dass positive Kooperationserfahrungen mit Personen, zu denen man nur schwache Assoziationen hat, den sogenannten *leap of faith* möglich machen, der notwendig ist, um Vertrauen über bestehende Gruppengrenzen hinweg zu erweitern. (Stolle (1998), S. 503f)

Die Wirksamkeit einer sozialen Vereinigung bei der Entwicklung generalisierter Vertrauenseinstellungen ist allerdings von den Ausprägungen verschiedener Parameter abhängig. Vorangestellt ist dabei die Frage, ob ein soziales Netzwerk eher *private social capital* oder *public social capital* produziert.

„Private social capital captures those attitudes and behaviors that mainly benefit the members of the voluntary association and group life itself. Such forms of social capital might be exclusive and not accessible to people outside of a particular group. Therefore, the use of private forms of social capital for situations outside of each group, that is, for situations of cooperation with people less known, is not obvious.“ (Stolle (2001a), S. 209f)

Dagegen beinhaltet *public social capital* "attitudes and behavior that go beyond the group life itself" (Stolle (2001a), S. 210).

Zunächst spielt die inhaltliche Ausrichtung der Vereinigung eine große Rolle. Zwar scheinen alle sozialen Netzwerke einen Einfluss auf die Entwicklung *generalisierten Vertrauens* zu haben, trotzdem tragen einige mehr bei als andere. So tragen insbesondere Organisationen, bei denen kulturelle Interessen, persönliche Interessen oder Interessen einer Gemeinschaft im Vordergrund stehen, zur Entwicklung *generalisierten Vertrauens* bei. Dagegen spielen Vereinigungen mit politischen oder ökonomischen Zielen eine geringere Rolle. (Stolle (2001a), S. 207, S. 234; Stolle/Rochon (1998), S. 55ff, S. 61; Stolle/Rochon (2001), S. 144, S. 148, S. 154)

Auch die Organisationsstruktur eines sozialen Netzwerkes wirkt sich auf sein Vermögen, *generalisiertes Vertrauen* zu produzieren, aus. STOLLE unterscheidet hier zwischen vertikalen und horizontalen Netzwerken. Vertikale Netzwerke weisen eine Asymmetrie in den sozialen Beziehungen auf, die das Erfahren von Gemeinsamkeit behindern kann. So fühlen sich Personen am unteren Ende der Hierarchie oft nur wenig integriert in Gruppenprozesse, was sich negativ auf *generalisiertes Vertrauen* auswirkt. Dagegen ermöglichen horizontale Netzwerke die Erfahrung von Wechselseitigkeit und Reziprozität. Je mehr Mitspracherecht die Mitglieder einer sozialen Vereinigung haben, desto größer deren Beitrag zur Generierung von *generalisiertem Vertrauen*. (Stolle (1998), S. 502; Stolle (2001b), S. 121; Stolle (2003), S. 26)

Ebenso hat die Art des Engagements der Gruppenmitglieder einen Einfluss auf *generalisiertes Vertrauen*. STOLLE differenziert hier Netzwerke, die *face-to-face*-Interaktionen fördern, von sogenannten „*checkbook*“ *organizations*, bei denen die Mitglieder lediglich in Form von Mitgliedsbeiträgen partizipieren. (Stolle (2003), S. 26) Ausgehend davon „that face-to-face contact with others is the key ingredient to the development of generalized trust“ (Stolle/Rochon (1999), S. 203), besitzen erstere ein größeres Gewicht bei der Generierung *generalisierten Vertrauens*.

Eine weitere Einflussgröße stellt die personelle Zusammensetzung des sozialen Netzwerkes dar. Heterogene Gruppen setzen sich aus Personen unterschiedlicher Herkunft zusammen und bieten die Möglichkeit, mit einer großen Auswahl an Vertretern verschiedener gesellschaftlicher Gruppen in Kontakt zu treten. Erfahrungen erfolgreicher Kooperation mit Personen, die sich von einem selbst unterscheiden, sind leichter auf eine ebenso heterogene Umwelt übertragbar. Diese Fähigkeit des „*bridging*“ hat einen positiven Einfluss auf die Entwicklung *generalisierten Vertrauens*. Demgegenüber stehen homogene Gruppen, deren Mitglieder sich in der Regel nur aus einem bestimmten limitierten Personenkreis der Gesellschaft rekrutieren. Hier ist lediglich eine Interaktion mit Personen möglich, die einem selbst sehr ähnlich sind. Solches „*bonding*“ wirkt sich entsprechend negativ auf die Generierung *generalisierten Vertrauens* aus. (Stolle (2001b), S. 122, S. 130; Stolle (2003), S. 26; Stolle/Hooghe (2003), S. 28; Stolle/Rochon (1998), S. 57ff; Stolle/Rochon (2001), S. 144, S. 152ff)

Neben diesen genannten positiven Effekten sozialer Vereinigungen auf die Generierung *generalisierten Vertrauens*, ist ihr Einfluss auch verschiedenen Einschränkungen unterworfen. So hat zwar die Mitgliedschaft an sich einen Effekt auf den Vertrauenslevel einer Person, allerdings lediglich in Form eines *trust boost* zu Beginn der Mitgliedschaft, der das neue Mitglied an den innerhalb der Vereinigung vorherrschenden Vertrauenslevel angleicht. Über die Zeit sind dagegen kaum Vertrauenszuwächse erkennbar, sodass die Länge der Mitgliedschaft keinen nachhaltigen Einfluss auf die Entwicklung *generalisierten Vertrauens* ausübt. (Stolle (1998), S. 510f, S. 516ff, S. 521; Stolle (2001a), S. 231ff; Stolle (2001b), S. 120, S. 130)

Darüber hinaus stellt sich die Frage, ob die Gruppenmitglieder *generalisiertes Vertrauen* erst innerhalb des jeweiligen sozialen Netzwerkes erlernen oder ob sie bereits bevor sie Mitglieder wurden, einen hohen Vertrauenslevel aufgewiesen haben. STOLLE findet in diesem Zusammenhang einen Selbstselektionseffekt, sodass sich vornehmlich Personen mit einem bereits hohen Vertrauenslevel für die Mitgliedschaft in einer sozialen Vereinigung entscheiden. (Stolle (1998), S. 507f, S. 521; Stolle (2001a), S. 229f, S. 233; Stolle (2001b), S. 130; Stolle (2003), S. 25)

Neben sozialen Netzwerken können allerdings auch noch andere Faktoren Auswirkungen auf die Generierung *generalisierten Vertrauens* haben. STOLLE nennt in diesem Zusammenhang Institutionen und den Staat, Quellen außerhalb des organisationalen Bereichs, wie Schule, Familie, Job und Gemeinschaft, Sekundäre Vereinigungen, wie beispielsweise soziale Bewegungen sowie informelle Kontakte, wie Freunde. STOLLE hebt dabei insbesondere die Rolle der Familie und die Rolle von staatlichen Institutionen hervor. (Stolle (1998), S. 500)

STOLLE geht davon aus, dass die familiäre Sozialisation einen der wichtigsten Faktoren bei der Entwicklung generalisierter Vertrauenseinstellungen darstellt. So fördert ein vertrauensvolles und offenes Elternhaus sowie eine Sozialisation in einer respektvollen und toleranten Atmosphäre die Vertrauensbereitschaft von Kindern. Eltern vermitteln darüber hinaus Bewertungsstandards hinsichtlich der Kooperationsbereitschaft und Vertrauenswürdigkeit Dritter sowie ihren Standpunkt hinsichtlich der Offenheit gegenüber Fremden. Außerdem fungiert die Familie als Lernarena für Kooperation. (Stolle (1998), S. 521f; Stolle (2001a), S. 235; Stolle (2001b), S. 131; Stolle (2003), S. 28ff)

Staatliche Institutionen wirken positiv auf die Entwicklung *generalisierten Vertrauens*, indem Sie ihrer Pflicht nachkommen, die soziale Ordnung zu bewahren und anomisches bzw. delinquentes Verhalten zu sanktionieren. Die Bürger übertragen dabei die guten bzw. schlechten Erfahrungen mit politischen und sozialen Institutionen auf die gesamte Gesellschaft. Es besteht dabei eine reziproke Beziehung zwischen dem Vertrauen in die Institutionen und *generalisiertem Vertrauen*, sodass sich ein kausaler Fluss von der Effizienz staatlicher Performanz über politisches Vertrauen hin zu generalisiertem sozialen Vertrauen ergibt. (Stolle (2003), S. 30ff Stolle (2004), S. 187f)

3.2.2.2 Culture-based Approach

Folgt man kultur-basierten Erklärungsansätzen, so fußt *generalisiertes Vertrauen* auf moralischen bzw. kulturellen Werten und Normen, die im Zuge der Sozialisation vermittelt werden und eine hohe Stabilität aufweisen. Es handelt sich hierbei um ein eher deterministisches Verständnis von der Genese *generalisierten Vertrauens*, spielen doch die aktuellen objektiven Lebensumstände nur eine untergeordnete Rolle. Vielmehr werden der Einfluss von Werten, die in frühen Lebensphasen erlernt wurden, und Idealen, die im späteren Leben übernommen wurden, sowie die aus deren Kombination resultierende Weltanschauung hervorgehoben. *Generalisiertes Vertrauen* spiegelt dann eine optimistische und egalitäre Weltsicht wider, die von den nachfolgenden Überzeugungen und Annahmen geprägt ist: (1) die Zukunft wird besser eingeschätzt, als die Vergangenheit gewesen ist, (2) Besitz von Kontrolle über unsere Umwelt und die Befähigung, diese formen zu können bzw. Kontrolle über das eigene Schicksal zu besitzen, (3) die Welt ist ein freundlicher Ort und die menschliche Natur ist grundlegend gut, (4) keine Angst davor zu haben, ausgenutzt oder betrogen zu werden.

Auch Zufriedenheit mit dem eigenen Leben bzw. ein ausgeprägtes subjektives Wohlbefinden sind Ausdruck dieses Optimismus gegenüber der Umwelt. Demgegenüber steht eine fatalistische Weltanschauung, die davon ausgeht, dass das Schicksal die Lebensbedingungen bestimmt, Erfolg durch Rahmenbedingungen außerhalb der individuellen Kontrolle bestimmt wird und Angst davor, ausgenutzt zu werden, gerechtfertigt ist. Verglichen mit zu Fatalismus neigenden Menschen weisen Personen, die eine optimistische und egalitäre Weltsicht verinnerlicht haben, eine höhere Wahrscheinlichkeit auf, Vertrauen auf Fremde zu erweitern, da sie auch unbekannte Personen als vertrauenswürdig annehmen. (Nannestad (2012), S. 425f; Park (2012), S. 4)

3.2.2.2.1 Vertrauen als Kontinuum mit zwei Polen (Uslaner)

USLANER unterscheidet Vertrauen in ein erfahrungsgeneriertes Vertrauen, welches sich aus Erlebnissen mit vertrauenswürdigen Personen ergibt, und ein kulturgeneriertes Vertrauen, welches einen vom Verhalten anderer weitestgehend unabhängigen individuellen Wert darstellt. (Uslaner (2008c), S. 726) Hierzu eröffnet er zunächst ein Vertrauensmodell bestehend aus einem Kontinuum zwischen *moralistic trust* und *strategic trust*. Während der Bezugspunkt von *strategic trust* das Individuum und seine Beziehungen zu spezifischen anderen Individuen ist, so bedeutet *moralistic trust* „not faith in specific people; rather, it is faith in the „generalized other““ (Uslaner (2002), S. 24).

Strategic trust beschreibt damit ein Vertrauen, das auf der individuellen Ebene wirksam wird und als solches, nach USLANERs Auffassung, mit Unsicherheit und Risiko verbunden ist. Unsicherheit und Risiko ergeben sich hierbei aus der Tatsache, dass sich diese Art des Vertrauens aus den Erfahrungen mit spezifischen Anderen generiert. Die Informationen, die uns dabei bezüglich einer Person zur Verfügung stehen, sind grundsätzlich mehr oder weniger beschränkt. So ist es prinzipiell nur dann möglich abzuschätzen, ob die Performanz eines Anderen in einem spezifischen Kontext den an ihn gesetzten Erwartungen entsprechen würde, wenn man über das Verhalten dieses Anderen in der fraglichen Situation Informationen besitzt. Das Verhalten in Situationen, über die keine Informationen vorliegen, kann nicht bestimmt werden. Stehen überhaupt keine Informationen über den Anderen zur Verfügung, kann sich demnach auch kein Vertrauen einstellen. *Strategic trust* wohnt darüber hinaus auch eine gewisse Fragilität inne, da neue Informationen hinsichtlich des Anderen ein zunächst positives Bild auch ins Negative verändern können, was ebenfalls einen Zusammenbruch der Vertrauensbeziehung zur Folge haben kann. (Uslaner (2002), S. 16f, S. 22, S. 24; Uslaner (2008a), S. 101ff; Uslaner (2008b), S. 290f)

Moralistic trust nimmt im Gegensatz zu *strategic trust* an, dass der Mensch an sich vertrauenswürdig ist. Es wird somit keine Aussage darübergemacht, ob ein spezieller Mensch dem in ihn gesetzten Vertrauen entsprechen würde oder ob er es missbrauchen würde. Vielmehr würde davon ausgegangen werden, dass er als Mensch prinzipiell vertrauenswürdig ist, selbst wenn er sich in einer spezifischen Situation bereits als vertrauensunwürdig erwiesen hätte. (Uslaner (2002), S. 17ff) *Moralistic trust* stellt somit ein „commandment to treat people as if they were trustworthy“ (Uslaner (2008a), S. 102) dar und ist in diesem Sinne „a moral dictate to treat others well, even in the absence of reciprocity“ (Uslaner (2002), S. 24). Die Grundlagen des *moralistic trust* werden bereits in den ersten Lebensjahren gelegt und verankern Vertrauen im Werte- und Normensystem eines Individuums. Besondere Eigenschaft solchen

3. Theorie: Vertrauen

Vertrauens ist seine weitestgehende Unempfindlichkeit für negative oder positive Einflüsse, sodass der von *moralistic trust* bereitgestellte Vertrauenslevel sehr stabil bleibt. (Uslaner (2008b), S. 290) Anderen Menschen wird dann implizit unterstellt, dass sie die gleichen Werte und Normen besitzen und befolgen und somit zur gleichen kulturellen Gemeinschaft gehören, wie man selbst. Hieraus ergibt sich folgende Definition bei USLANER:

„Moralistic trust is the belief that others share your fundamental moral values and therefore should be treated as you would be treated by them. The values they share may vary from person to person. What matters is a sense of connection with others because you see them as members of your community whose interests must be taken seriously. (...) When others share our basic premises, we face fewer risks when we seek agreement on collective action problems.“ (Uslaner (2002), S. 18)

Sowohl *strategic trust* als auch *moralistic trust* fördern Kooperation, indem sie Transaktionskosten verringern. *Strategic trust* wird dabei allerdings nur auf der Individualebene wirksam und kann lediglich Kooperationen zwischen Personen herbeiführen, die sich wechselseitig bekannt sind. Kooperationen auf der Kollektivebene werden durch *moralistic trust* gefördert, da an dieser Stelle eine Interaktion zwischen Fremden, d.h. Personen, die wechselseitig keinerlei Informationen übereinander besitzen, etabliert werden muss. Hierbei ist die von *moralistic trust* bereitgestellte grundsätzlich vertrauensvolle Einstellung notwendig, da nicht von Erfahrungen mit bekannten Personen auf das Verhalten von Fremden geschlossen werden kann. *Moralistic trust* ermöglicht damit auch eine Erweiterung der Opportunitäten eines Individuums. (Uslaner (2002), S. 17, S. 20f; Uslaner (2008b), S. 290f)

„This type of trust in strangers is an essential foundation of a civil society. I call it ‚moralistic trust‘. This is trust in people whom we don’t know and who are likely to be different from ourselves. We can’t base trust in strangers on their trustworthiness, because there is no way for us to know whether they are honorable. Instead, we must presume that other people are honorable. We believe that others share our fundamental moral values. Moralistic trust provides the rationale for getting involved with other people and working toward compromises.“ (Uslaner (2002), S.15)

USLANER kommt bezüglich der Differenz von *strategic trust* und *moralistic trust* zu folgender Schlussfolgerung:

„Strategic trust reflects our expectations about how people *will* behave. Moralistic trust is a statement about how people *should* behave. *People ought to trust each other.*“ (Uslaner (2008a), S. 103)

In die Unterscheidung von *strategic* und *moralistic trust* eingebettet findet sich ein weiteres Kontinuum mit den Polen *generalized trust* und *particularized trust*. Während *generalized trust* an *moralistic trust* angelagert ist und dieses in seiner Universalität beschneidet, gleichzeitig aber um eine erfahrungsbasierte Komponente erweitert, ist *particularized trust* an *strategic trust* angelagert und erweitert dieses um die Anwendung von Erfahrungen mit speziellen Individuen auf Gruppen, die aus ähnlichen Individuen bestehen.

Particularized trust beschreibt USLANER folgendermaßen:

„Particularized trust is the notion that we should only have faith in people like ourselves, and this restricts the size of our moral community. We think that we know about people like ourselves, so particularized trust is more likely to reflect our experience.“ (Uslaner (2002), S. 21)

3. Theorie: Vertrauen

Particularized trust beschränkt Vertrauen demnach auf Personen, die zur *ingroup* des Vertrauenden zählen, wobei die Vergabe von Vertrauen bzw. Misstrauen gleichermaßen auf Erfahrung mit vertrauenswürdigen bzw. -unwürdigen Personen oder Stereotypen basiert. Die Umwelt wird vom Vertrauenden als Bedrohung aufgefasst, über die man nur wenig Kontrolle hat, was einer weiteren Segregation des Vertrauenden von der Umwelt Vorschub leistet. Die *outgroup* wird dabei vom Vertrauenden nicht als Teil seiner kulturellen Gemeinschaft wahrgenommen, was auch zur Folge hat, dass Vertrauensbrüche gegenüber Mitgliedern dieser Gruppe als weniger schwerwiegend angesehen werden. (Uslaner (2002), S. 27f, S. 31; Uslaner (2008a), S. 104; Uslaner (2008b), S. 293; Uslaner (2012), S. 7)

Generalized trust basiert auf *moralistic trust*, wobei sich der Unterschied zwischen beiden im Wesentlichen daraus ergibt, dass letzteres einer Art von Gebot gleichkommt, während ersteres ein Ausdruck des Ausmaßes der Verinnerlichung dieses Gebotes darstellt. Das Ausmaß von *generalized trust* resultiert dabei aus der Stärke einer optimistischen Grundhaltung gegenüber anderen Menschen und der Umwelt sowie der Stärke eines Gefühls der Kontrolle über das eigene Schicksal. (Uslaner (2002), S. 21, S. 26ff; Uslaner (2004), S. 502; Uslaner (2008a), S. 103f, S. 108ff; Uslaner (2008b), S. 293)

„Generalized trust is the perception that most people are part of your moral community. Its foundation lies in moralistic trust but it is not the same thing. Generalized trust is a measure of the scope of our community, and it is based upon both morals and our collective experiences. Sometimes things look good and sometimes they don't. Our values (moralistic trust) don't change readily. But the way we interpret them does reflect some experiences from daily life. And this is what distinguishes generalized trust from moralistic trust: Generalized trust goes up and down, though it is basically stable. Moralistic trust is a more lasting value.“ (Uslaner (2002), S. 26)

Generalized trust beschreibt „the belief that ‚most people can be trusted‘ (...)“ (Uslaner (2012a), S. 7). Dies schließt nicht mehr nur Personen der *ingroup* ein, vielmehr wird es durch diese Art des Vertrauens möglich, auch Fremden zu vertrauen. Neben der *ingroup* werden nun auch Fremde als Teil der eigenen kulturellen Gemeinschaft betrachtet. Es ergibt sich auf diese Weise die Möglichkeit, Kooperationen zwischen Personen aufzubauen, die vollkommen verschieden voneinander sind. (Uslaner (2004), S. 501; Uslaner (2008a), S. 103f; Uslaner (2012a), S. 7f)

Abschließend lässt sich die Differenz zwischen *particularized trust* und *generalized trust* wie folgt zusammenfassen:

„Particularized trusters only rely upon people they are sure share their own values. Generalized trusters presume that most people they meet share their values; particularized trusters demand evidence that people outside their own circles (or identity groups) share their beliefs.“ (Uslaner (2002), S. 28)

Ausgehend von den bisherigen Erörterungen lässt sich erkennen, dass USLANER *generalisiertes Vertrauen* als kulturellen Wert im Sinne einer moralischen Verpflichtung betrachtet. *Generalisiertes Vertrauen* stellt hiernach eine stabile individuelle Prädisposition dar, die durch eine optimistische Weltanschauung, eine Überzeugung, Herr über das eigene Schicksal zu sein, und die Annahme, dass die meisten Menschen Teil der gleichen Wertegemeinschaft sind und man deshalb ein gemeinsames Schicksal sowie fundamentale moralische Werte miteinander teilt, gekennzeichnet ist. (Uslaner (2001), S. 106; Uslaner (2002), S. 76ff; Uslaner (2005), S. 76f; Uslaner (2006), S. 147f; Uslaner (2008c), S. 726, S. 733; Uslaner (2013), S. 629ff)

3. Theorie: Vertrauen

Als Determinanten *generalisierten Vertrauens* bestimmt USLANER die familiäre Sozialisation, religiöse Werte, Bildung, ethnische Diversität sowie die Ressourcen- und Chancenverteilung bzw. das Ausmaß der sozioökonomischen Ungleichheit innerhalb einer Gesellschaft.

Die Sozialisation in einer vertrauensvollen Familie, innerhalb derer man ein gutes Verhältnis zu seinen Eltern hatte, führt in der Regel zur Entwicklung *generalisierten Vertrauens* im Erwachsenenalter. Vertrauensvolle, tolerante und sozialintegrierte Eltern fungieren hierbei als vorbildhaftes Modell für ihre Kinder. Die generalisierte Vertrauenseinstellung kann dabei sogar auf einer mehrere Generationen übergreifende Historie zurückgeführt werden, wobei auch das Herkunftsland der Vorfahren eine große Rolle spielt. (Uslaner (2002), S. 92f, S. 110; Uslaner (2008a), S. 109; Uslaner (2008c), S. 726ff; Uslaner (2012a), S. 9; Uslaner (2013), S. 631)

Religiöser Glaube spielt ebenso eine nicht zu verachtende, allerdings auch zwiespältige Rolle bei der Entwicklung generalisierter Vertrauenseinstellungen. Er formt die Weltanschauung des Individuums und sorgt für eine Abkehr vom puren Selbstinteresse hin zu einer Gemeinschaftsorientierung. Gleichzeitig besteht aber auch die Möglichkeit, dass religiöse Werte ein Misstrauen gegenüber denjenigen hervorrufen, die einen Glauben besitzen, der vom eigenen abweicht. (Uslaner (2001), S. 104f; Uslaner (2002), S. 86ff; Uslaner (2008c), S. 729, S. 733)

Ethnische Diversität innerhalb einer Gesellschaft hat sowohl auf die Entwicklung von partikularem als auch von *generalisiertem Vertrauen* einen Einfluss. So wirkt das Leben innerhalb einer Gesellschaft, die eine hohe ethnische Homogenität aufweist, positiv auf die Entwicklung *partikularisierten Vertrauens*. In einer sozialen Umgebung, die sich ausschließlich aus Personen zusammensetzt, die einem selbst sehr ähnlich sind, wird für Kooperationen jeglicher Art kein *generalisiertes Vertrauen* benötigt. Dagegen ermöglicht eine stark ausgeprägte ethnische Diversität innerhalb einer Gesellschaft Kontakt zu einer Vielzahl von Menschen, die sich von einem selbst unterscheiden, was die Entwicklung *generalisierten Vertrauens* fördert. Kooperation mit Migranten, Minderheiten bzw. im Allgemeinen mit Personen, die sich von einem selbst abheben, bewirkt, dass diese Personen eher als Teil der eigenen Wertegemeinschaft angesehen werden. (Uslaner (2008a), S. 115; Uslaner (2012a), S. 6ff; Uslaner (2013), S. 632)

Eine gerechte Verteilung der sozioökonomischen Ressourcen innerhalb einer Gesellschaft trägt zur Entwicklung sozialer Solidarität und sozialer Kohäsion bei. In der Folge entwickelt sich die Überzeugung, dass alle Menschen ein gemeinsames Schicksal und gleiche grundlegende Werte miteinander teilen. Diese Erfahrung von Gemeinsamkeit lässt die Menschen sich als Teil einer größeren sozialen Ordnung begreifen, was wiederum die Entwicklung und Verbreitung *generalisierten Vertrauens* fördert. Darüber hinaus bewirkt eine sinkende soziale Ungleichheit, dass die Menschen optimistischer in die Zukunft blicken. Der damit einhergehende Relevanzverlust von speziellen Beziehungen und besonderen Netzwerken für das eigene Fortkommen, nährt außerdem die Überzeugung, das eigene Schicksal kontrollieren zu können. Beides stellt die Voraussetzung für *generalisiertes Vertrauen* dar. (Rothstein/Uslaner (2005), S. 42ff, S. 61, S. 51f, S. 69; Uslaner (2003), S. 179ff; Uslaner (2008a), S. 113; Uslaner (2008b), S. 291; Uslaner/Brown (2005), S. 870)

3.2.3 *Institution-centered Accounts*

Institutionen-zentrierte Ansätze gehen davon aus, dass die Quelle *generalisierten Vertrauens* im Verhalten staatlicher Institutionen zu suchen ist. So wird angenommen, dass politische Institutionen die Menge und den Typus sozialen Kapitals innerhalb einer Gesellschaft gleichermaßen erschaffen, kanalisieren und beeinflussen können und auf diese Weise indirekt mittels der Manipulation der Zivilgesellschaft auf *generalisiertes Vertrauen* wirken. Damit Soziales Kapital und in der Folge auch *generalisiertes Vertrauen* anwächst, müssen diese somit in einen Kontext aus politischen, administrativen und judikativen Institutionen eingebettet sein bzw. mit diesem verbunden sein. Eine besondere Rolle für die Genese *generalisierten Vertrauens* spielen die durch die Institutionen gesetzten und in verschiedenen Gestalten auftretenden Anreize, sich vertrauenswürdig zu verhalten. (Nannestad (2008), S. 423; Park (2012), S. 5; Rothstein/Stolle (2002), S. 7f; Rothstein/Stolle (2007), S. 6f; Rothstein/Stolle (2008a), S. 278f; Rothstein/Stolle (2008b), S. 443)

Zunächst gestalten und etablieren politische Institutionen normative Standards und setzen diese gegenüber den Bürgern durch. Das Ausmaß *generalisierten Vertrauens* wird dabei umso größer, je mehr Bürger auf die Einhaltung dieser Normen verpflichtet werden können. Abgesichert wird diese Verpflichtung durch Möglichkeiten der Sanktionierung bei Vertrauensmissbrauch, woraus auch eine Verminderung des mit dem Schenken von Vertrauen verbundenen Risikos resultiert. Von besonderer Relevanz sind in diesem Zusammenhang institutionelle Fairness und Unparteilichkeit bzw. Objektivität, werden doch allen Bürgern gleiche Rechte und Pflichten zugesprochen. Bevorzugung bestimmter sozialer Gruppen oder Willkür bei der institutionellen Durchsetzung normativer Regeln würde zu einer Unterminierung *generalisierten Vertrauens* führen. Politische Institutionen erschaffen somit gesellschaftliche Rahmenbedingungen, die förderlich für die Entwicklung *generalisierten Vertrauens* sind. (Nannestad (2008), S. 424; Park (2012), S. 5)

Darüber hinaus weisen Institutionen eine Vorbildfunktion auf, die sich darin manifestiert, dass von ihren Institutionenvertretern auf die übrigen Gesellschaftsmitglieder geschlossen wird bzw. von deren Verhalten universelle und adäquate Handlungsmaximen abgeleitet werden. So wird aus dem Fehlverhalten einzelner Beamter mitunter auf sämtliche Vertreter von einer Institution geschlossen oder sogar auf Menschen im Allgemeinen. In der Folge würde man sich selbst gezwungen sehen, moralisch falsch zu handeln, um keinen Wettbewerbsnachteil zu erleiden, was negative Auswirkungen auf die Verbreitung *generalisierten Vertrauens* hätte. Erfahrungen mit Institutionen, die ihr Handeln an Wahrheits- und Gerechtigkeitsnormen orientieren, die Korruption vermeiden, Solidarität und Chancengleichheit verpflichtet sind sowie Vertrags- bzw. Rechtsbrüche sanktionieren, können dagegen in Vertrauen gegenüber Fremden umgesetzt werden. (Nannestad (2008), S. 424; Park (2012), S. 5)

Nicht zuletzt können Institutionen auch als Substitute für Vertrauen fungieren. Hierbei spielt einerseits die wechselseitige Beeinflussung von politischem Vertrauen und *generalisiertem Vertrauen* eine Rolle. Werden staatliche Institutionen als vertrauenswürdig angesehen, so hat dies positive Auswirkungen auf *generalisiertes Vertrauen*. Andererseits kann der vollständige Verlass auf die von den Institutionen vertretenen und realisierten normativen Standards als Ersatz für Vertrauen dienen. Vertrauen wird in diesem Fall durch Recht ersetzt. (Nannestad (2008), S. 424; Park (2012), S. 5;)

Grundlegende Kritik an den institutionen-zentrierten Ansätzen entzündet sich daran, dass es sich bei *generalisiertem Vertrauen* um eine über lange Zeit stabile und sich nur langsam

verändernde Form des Vertrauens handelt. Als solches wird es nur wenig durch unmittelbare Erfahrungen beeinflusst, was den Effekt von positiven Erlebnissen mit Institutionen auf *generalisiertes Vertrauen* fraglich macht. Vielmehr müsste es sich bei dem durch Institutionen beeinflussten Vertrauen um eine Form *partikularen Vertrauens* handeln. (Uslaner (2002), S. 44)

3.2.3.1 Attitudinal Approaches

Im Rahmen einstellungsbezogener Ansätze wird die Beziehung zwischen Institutionenvertrauen bzw. politischem Vertrauen und *generalisiertem Vertrauen* betrachtet, wobei die kausale Richtung des Effekts unklar bleibt und derzeit von einer wechselseitigen Beeinflussung ausgegangen wird. Nimmt man *generalisiertes Vertrauen* als Prädiktor für politisches Vertrauen an, so geht man davon aus, dass sich Institutionenvertrauen aus einer optimistischen Grundhaltung und Zufriedenheit der Bürger ergibt. Tatsächlich scheint allerdings der Einfluss von Institutionenvertrauen auf *generalisiertes Vertrauen* größer zu sein als andersherum. (Rothstein/Stolle (2002), S. 8f; Rothstein/Stolle (2007), S. 8; Rothstein/Stolle (2008a), S. 279f; Rothstein/Stolle (2008b), S. 443)

Einstellungsbezogene Ansätze sehen sich einer Vielzahl von Kritikpunkten gegenüber. Zunächst bleibt unbestimmt, ob die auf Institutionen bezogenen Einstellungen unabhängig sind von deren tatsächlichen institutionellen Charakteristika. So muss angenommen werden, dass eine Bewertung von Institutionen nicht losgelöst von deren Eigenarten oder deren Performanz erfolgt. Dies widerspricht allerdings der Dauerhaftigkeit und Robustheit *generalisierten Vertrauens*, müsste dieses doch dann je nach Gunst für die Institutionen mehr oder weniger stark schwanken. (Rothstein/Stolle (2002), S. 9f; Rothstein/Stolle (2007), S. 9; Rothstein/Stolle (2008a), S. 280; Rothstein/Stolle (2008b), S. 443f)

Des Weiteren existieren verschiedene Formen institutionellen Vertrauens, die unter ein einziges Label subsummiert werden. Hierbei werden auch Institutionen einbezogen, die nur einen geringen Einfluss auf *generalisiertes Vertrauen* haben. Die Folge sind schwache bis keine Korrelationen zwischen Institutionenvertrauen und *generalisiertem Vertrauen*. (Rothstein/Stolle (2002), S. 9f; Rothstein/Stolle (2007), S. 9; Rothstein/Stolle (2008a), S. 280; Rothstein/Stolle (2008b), S. 443f)

Zuletzt bleibt unklar, welcher kausale Mechanismus die Verbindung von Institutionenvertrauen und *generalisiertem Vertrauen* bestimmt. Im Einzelnen existiert kein Verständnis dafür, wie Personen, die ein hohes Niveau an *generalisiertem Vertrauen* aufweisen, einen positiven Beitrag zur Performanz von politischen Institutionen leisten bzw. weshalb sie ein positiveres Bild von politischen Institutionen besitzen. (Rothstein/Stolle (2002), S. 9f; Rothstein/Stolle (2007), S. 9; Rothstein/Stolle (2008a), S. 280; Rothstein/Stolle (2008b), S. 443f)

3.2.3.1.1 Staat und Vertrauen (Levi)

LEVI begreift Vertrauen als ein Konzept, das sowohl eine grundlegende Relationalität als auch eine grundlegende Konditionalität aufweist.

„Trust is relational, involving at least one individual making themselves vulnerable to another individual, set of individuals, or institutions that possess the potential to do harm or betray. Trust implies a judgement of risk in conditions of uncertainty. Moreover, trust is seldom unconditional; it is given over specific domains.“ (Levi (2001), S. 15922)

Vertrauen ist somit immer auf einen bestimmten Akteur hinsichtlich eines spezifischen Gegenstandes gerichtet. Ob Vertrauen vergeben wird, hängt dabei im Wesentlichen von der Vertrauenswürdigkeit des Vertrauten bzw. der Vertrauenswürdigkeit der Quelle, aus der man Informationen über den Vertrauten schöpft, ab. Vertrauenswürdigkeit setzt sich dabei aus zwei Komponenten zusammen: (a) der Verpflichtung des Vertrauenden im Interesse des Vertrauenden zu handeln und (b) der Kompetenz des Vertrauten bezüglich der Vertrauenssache. (Levi (2000), S. 139; Levi (2001), S. 15922f; Levi/Stoker (2000), S. 476)

Nach LEVI kann sich die Motivation zu vertrauen aus vier verschiedenen Quellen ergeben: (1) Als Folge einer Disposition, die einige Menschen per se vertrauensvoller bzw. vertrauenswürdiger macht als andere. Sie kann angeboren sein oder sich aus Erfahrungen in der Kindheit, mit sozialen Netzwerken und Bindungen oder mit sozialen Normen entwickeln. (2) Als Folge von Moralität, wonach Vertrauen eine Tugend darstellt und Ausdruck eines guten Charakters ist. (3) Als Folge von Sozialität und geteilter Identität, sodass solchen Personen getraut wird, die ähnliche Überzeugungen wie man selbst teilt und einen in seinen Überzeugungen bestätigen und unterstützen. (4) Als Folge einer Kalkulation, die Vertrauen als lohnenswert kennzeichnet und das mit der Gewährung von Vertrauen verbundene notwendige Risiko rechtfertigt. (Levi (2001), S. 15923f)

Es ergeben sich dann zwei grundsätzliche Ebenen der Konzeptualisierung von Vertrauen. So kann Vertrauen entweder kognitiver oder nichtkognitiver Natur und gleichzeitig entweder Bestandteil dyadischer Beziehungen oder in soziale Netzwerke und institutionelle Arrangements eingebettet sein. Nichtkognitives Vertrauen beruht auf der Attribution von Eigenschaften die Vertrauenswürdigkeit anzeigen auf Personen und Institutionen. Grundlage hierfür können neben sozialen Bindungen und Gruppenidentifikation auch Emotionen und moralische Überzeugungen sein. Kognitives Vertrauen benötigt eine Abschätzung des Vertrauensrisikos und eine rational begründete Erwartung, dass der Vertraute vertrauenswürdig sein wird. Basis solchen Vertrauens sind meist Informationen über das vergangene oder gegenwärtige Verhalten des Vertrauten. Die Ebene dyadischen und eingebetteten Vertrauens verläuft quer zur Ebene kognitiven bzw. nichtkognitiven Vertrauens, sodass eine Kombination beider Ebenen möglich wird. Während dyadisches Vertrauen an konkrete Andere gebunden ist und auf der Individualebene verläuft, basiert eingebettetes Vertrauen auf sozialen Rollen, Kategorien oder Regeln und beinhaltet Konzepte wie *generalisiertes Vertrauen*. (Levi (2001), S. 15924)

LEVI geht davon aus, dass „Trust is not one thing and does not have one source; it has a variety of forms and causes“ (Levi (1998), S. 79) und dass die Produktion von Vertrauen in Gesellschaften dementsprechend durch verschiedene Quellen geleistet werden kann. Infrage kommen dabei sämtliche Gelegenheiten, die es ermöglichen, die Notwendigkeit von persönlicher Involviertheit in die Suche von Informationen über potentielle Vertraute zu verringern, wie beispielsweise die individuelle Fähigkeit zum Charakterassessment, Netzwerke oder Institutionen. Letztendlich spielen nach LEVIs Überzeugung allerdings Institutionen hierbei die gewichtigste Rolle, bieten sie doch eine Form der Rückversicherung für den Vertrauenden. (Levi (1998), S. 79)

Netzwerke können zur Etablierung von *generalisiertem Vertrauen* bzw. Vertrauen gegenüber Fremden beitragen, indem sie Heuristiken der Vertrauensvergabe verbreiten. Die Mitgliedschaft innerhalb eines Netzwerkes lehrt eine Person nun, ihre eigene Vertrauenswürdigkeit auf andere Personen zu projizieren und diese entsprechend der eigenen Selbstwahrnehmung als ebenfalls vertrauenswürdig anzusehen. Insofern alle anderen Gesellschaftsmitglieder in ähnlicher Art und

3. Theorie: Vertrauen

Weise sozialisiert wurden und in gleicher Art und Weise handeln, lässt sich Vertrauen über die gesamte Gesellschaft generalisieren. LEVI schränkt allerdings ein, dass eine Vielzahl von verschiedenartigen sozialen Netzwerken, Organisationen und Verbindungen existiert, die mehr oder weniger nützlich bzw. zielführend für die Vertrauensproduktion sein können. Zusätzlich verhält es sich oft so, dass eine innerhalb eines Netzwerkes entwickelte Vertrauensnorm nur selten auf Individuen außerhalb dieser Vereinigung übertragen wird. Vielmehr ergibt sich meist ein gegenteiliger Effekt, sodass eine Abschottung der *ingroup* gegenüber der *outgroup* resultiert und Vertrauen einzig den Mitgliedern der *ingroup* entgegengebracht wird. (Levi (1996), S. 46ff)

Wie schon weiter oben beschrieben, hält LEVI den Staat bzw. Institutionen für die größten Produzenten *generalisierten Vertrauens*. So stellt sich *generalisiertes Vertrauen* in der Regel auf Basis deren Fähigkeit, die Rechtsordnung zu überwachen und zu gewährleisten, ein. So werden beispielsweise Verstöße sanktioniert. Daneben spielt auch die Bereitstellung von Informationen und Garantien bezüglich Individuen die vertrauenssuchend sind, eine große Rolle. (Levi (1998), S. 87) Damit staatliche Institutionen dieser Rolle gerecht werden können, sind sie darauf angewiesen, dass ihnen ihrerseits Vertrauen entgegengebracht wird.

„If the state is one of the institutions – and, in many cases, the most important institution – for promoting generalized trust, it can play this role only if the recipients of these services consider the state itself to be trustworthy.“ (Levi (1998), S. 86)

Dieses Vertrauen in staatliche Institutionen kann auf verschiedenen Wegen realisiert werden (Levi (1998), S. 86ff):

- (1) *Vertrauenswürdige politische Akteure*: Staatliche Akteure sollten sich konsistent vertrauenswürdig verhalten, indem sie prinzipiengebunden und ohne Selbstinteresse handeln.
- (2) *Glaubhafte Zusagen staatlicher Akteure*: Bindung der staatlichen Akteure an geschlossene Kompromisse und bestimmte Standards, sodass die Notwendigkeit einer Überwachung des Staates durch den Bürger reduziert wird.
- (3) *Faire Prozeduren*: Institutionen sollten sich grundsätzlich fair und unvoreingenommen gegenüber den Bürgern verhalten.
 - (a) *Zwang*: Sicherstellung der Befolgung staatlicher Vorgaben und dass jeder seinen Beitrag leistet.
 - (b) *Universalismus*: Anwendung allgemeingültiger Kriterien bei der Rekrutierung und Förderung staatlicher Vertreter sowie der Regulierung staatlicher und zivilgesellschaftlicher Institutionen.
 - (c) *Unvoreingenommene Institutionen*: Fairness und Gerechtigkeit von Entscheidungsprozessen.
 - (d) *Partizipation*: Einbezug der Bürger in den politischen Prozess.
- (4) *Reziprozität*: Wahrnehmung der Bürger für ihr Vertrauen in den Staat eine Gegenleistung zu erhalten.

3.2.3.1.2 Vertrauen als Folge sozialer Ordnung (Offe)

OFFE sieht Vertrauen in letzter Instanz mit dem Problem der Aufrechterhaltung sozialer Ordnung verknüpft. Ausgangspunkt dieser Überlegung ist die allgemeine Annahme, dass soziale Ordnung in einer modernen Gesellschaft auf drei grundlegenden Koordinationsmedien und

3. Theorie: Vertrauen

deren Zusammenwirken basiert: *Geld*, *Herrschaft* und *Wissen*. *Geld* regelt dabei das Handeln innerhalb der ökonomischen Sphäre, während *Herrschaft* mittels legitimer Gewalt das gesellschaftliche Handeln reguliert. Durch *Wissen* wird schließlich vermittelt, welchen Problemen sich eine moderne Gesellschaft gegenüber sieht und welche Strategien zu ihrer Lösung angewandt werden können. OFFE sieht die Etablierung sozialer Ordnung, die ausschließlich auf der Kooperation dieser drei Koordinationsmedien basiert, als unvollkommen an, werden hierbei doch nur manifeste Determinanten des Sozialen für eine Erklärung des Ordnungsproblems herangezogen. „Informelle Modi der sozialen Koordination“ (Offe (2001a), S. 242) bleiben nur unzureichend berücksichtigt oder unberücksichtigt.

„Diese informellen Modi der sozialen Koordination beruhen auf Bindungen, die auf lebensweltlich generierten Vorstellungen und Erwartungen gründen, die Menschen moderner Gesellschaften von anderen Menschen in diesen Gesellschaften haben, und nach denen diese Menschen handeln.“ (Offe (2001a), S.242)

Selbst wenn die drei genannten manifesten Koordinationsmedien vollkommen integriert sind, ist deren Koordination immer noch von der Bereitschaft zur Kooperation abhängig. Vertrauen scheint einer der Schlüssel zu dieser Kooperationsbereitschaft zu sein¹¹. (Offe (2001, S. 241ff)) Allerdings besteht zwischen den manifesten Kontrollmedien und Vertrauen auch eine positive Rückkopplung. Je mehr Verfügungsgewalt man über Geld, Macht und Wissen hat, desto geringer wiegt ein Vertrauensbruch und desto leichter ist es Vertrauen zu schenken. Die Konsequenzen, die sich aus dem Scheitern einer Vertrauensbeziehung ergeben, können mit Hilfe der manifesten Kontrollmedien kompensiert werden. Außerdem ist es mit Hilfe der manifesten Kontrollmedien möglich Zwänge zu errichten, was bedeutet:

„Sollte sich jemand meinen Wünschen und Erwartungen widersetzen, kann ich dessen Willen brechen und sie zum Gehorsam zwingen“ (Offe (2001a), S. 259).

Nachteil eines solchen Vorgehens ist, dass die positiven Effekte von Vertrauensbeziehungen, nämlich Innovationen, nicht eintreten, da dem Vertrauensnehmer keinerlei Handlungsspielraum eingeräumt wird. (Offe (2001a), S. 258ff) Vertrauen scheint damit ein Konstrukt zu sein, das dem *Matthew-Effekt*¹² unterliegt:

„Das Paradox liegt also darin, dass diejenigen, die am meisten auf Vertrauen angewiesen wären (weil ihnen sonst nichts zur Verfügung steht), das eingegangene Risiko am wenigsten tragen können, während diejenigen, die es am wenigsten benötigen, sich Vertrauen am leichtesten leisten und daraus Vorteile erzielen können.“ (Offe (2001a), S. 260)

¹¹ Besondere Qualität informeller Koordinationsmedien ist, dass sie in der Regel nicht durch spezifisches zielgerichtetes Handeln aktiv herbeigeführt werden können. Sie sind entweder vorhanden oder nicht. OFFE verweist in diesem Zusammenhang auf die (Un)Wirksamkeit vertrauensbildender Maßnahmen: „Wenn sie von der ‚anderen‘ Seite nicht als vertrauenswürdig, sondern als trügerische Inszenierung wahrgenommen werden, die nur strategischen Interessen dienen, werden sie nicht Vertrauen, sondern gerade im Gegenteil Zynismus hervorrufen. (...) Untaugliche Versuche zur Vertrauensbildung können Misstrauen hervorrufen und so zum Boomerang werden. Das Vertrauen in die Vertrauenswürdigkeit der Urheber vertrauensbildender Maßnahmen muss schon vorhanden sein, wenn diese mit ihren Maßnahmen zum Erfolg kommen sollen.“ (Offe (2001a), S. 243f) An anderer Stelle heißt es außerdem: „Wer Teile der Bevölkerung zum Objekt ‚vertrauensbildender Maßnahmen‘ erwählt, provoziert damit womöglich gerade den Argwohn und das Misstrauen der Adressaten.“ (Offe (2000), S. 129) Des Weiteren erzeugt eine vom Adressaten als übervorteilend wahrgenommene vertrauensbildende Maßnahme keine Reziprozitätspflicht, was einem Vertrauensmissbrauch Vorschub leisten kann. (Offe (2001a), S. 257)

¹² Vgl. Abschnitt „3.1.1.4.2.2 Der Treuhänder“

3. Theorie: Vertrauen

Vertrauen scheint somit tatsächlich ein übergeordnetes Steuerungsmedium zu sein, das auf verschiedenen Ebenen zur Etablierung sozialer Ordnung beiträgt. OFFE unterscheidet hier (1) Bürger - Mitbürger, (2) Bürger - politische/sectorale Eliten (z.B. Repräsentanten von Kirche, Medien, Militär, Polizei, Gerichten etc.) (3) politische Eliten - sectorale Eliten und (4) Eliten - Bürger bzw. die Öffentlichkeit insgesamt. Sein Interesse gilt hierbei insbesondere den Vertrauensbeziehungen zwischen normalen Bürgern (Offe (2001a), S. 244)

„Vertrauen ist eine Überzeugung über das Handeln des anderen. Diese Überzeugung bezieht sich auf die Wahrscheinlichkeit, dass bestimmte Gruppen von anderen in einer bestimmten Weise handeln oder nicht handeln werden, wobei diese Handlungen/Unterlassungen das Wohlergehen dessen, der vertraut (oder einer für ihn relevanten Gruppe) betreffen. Vertrauen ist demnach die kognitive Erwartung, dass andere durch ihr Handeln oder Unterlassen zum Wohlergehen eines einzelnen oder einer Gruppe beitragen, jedenfalls von schädigenden Handlungen absehen.“ (Offe (2001a), S. 249)

OFFE unterscheidet ähnlich wie LUHMANN und GIDDENS zwischen Vertrauen (*trust*) und „subjektiver Gewissheit“ (*confidence*)“ (Offe (2001a), S. 244). Vertrauen bezieht sich dabei auf ein bewusstes Handeln gegenüber anderen Personen. Das heißt: Vertrauen wohnt eine gewisse Verantwortlichkeit für das eigene Verhalten inne. Dagegen unterliegt Gewissheit keiner bewussten Einflussnahme. Entsprechend gilt:

„Gewissheit‘ verhält sich zu Vertrauen, wie sich Tatsachen zu Handlungen verhalten.“ (Offe (2001a), S. 244)

Konsequenzen, die sich aus Vertrauen ergeben, werden demnach auf das vertrauende Individuum attribuiert, während Konsequenzen aus Gewissheit dem Zufall zugeschrieben werden.¹³ Vertrauen beinhaltet somit eine reflexive Komponente, der die Gewissheit fehlt.

Wer vertraut, geht immer ein doppeltes Risiko ein bzw. ist mit Kontingenz konfrontiert. Die Überzeugungen, die zu einem vertrauensvollen Handeln führen, erfahren keine Letztbegründung, sondern sind lediglich als Antizipationen des Handelns des Gegenübers zu werten. Entsprechend besteht bei jeder Vertrauensvergabe das Risiko einer falschen Annahme über das zukünftige Handeln des Gegenübers. Direkt daran anschließend ergibt sich das Risiko des Verlustes bei einem Vertrauensmissbrauch. Es ist für den Vertrauenden unmöglich vorherzubestimmen, ob der, dem man vertraut, auch im Sinne der Erwartungen an ihn handeln wird. (Offe (2001a), S. 248ff)

Die hier beschriebenen Risiken potenzieren sich im Falle von leichtfertigem bzw. blindem Vertrauen. In einem solchen Fall erhöhen sich die Risiken durch eine breite Streuung des Vertrauens und das niedrige Reflexionslevel bei der Vergabe des Vertrauens. Allerdings besteht die Lösung des Risikoproblems nicht in der Vermeidung dieser Risiken: So ist es eine besondere Eigenschaft von Vertrauen, die Handlungsoptionen des Vertrauenden zu erweitern. Dies geschieht durch die Ausdehnung des vertrauten sozialen Umfeldes um neue Kooperationspartner. Der Vertrauende geht beim Verzicht auf die Vergabe von Vertrauen also das Risiko ein, die potentiellen Gewinne aus einer neuen Vertrauensbeziehung ziehen zu lassen. Die Vorteile aus solchen neuen Vertrauensbeziehungen überwiegen auch dann, wenn die

¹³ Nach OFFEs Auffassung impliziert die Unterscheidung von Vertrauen (*trust*) und Gewissheit (*confidence*) eine Abkehr vom Begriff des Institutionenvertrauens. Institutionen und die von ihnen verkörperten Normen weisen eine besondere zeitliche Beständigkeit auf, sodass ihrer Gültigkeit mit einer „zuversichtlichen Gewissheit“ begegnet werden kann. Vertrauen in Institutionen scheint sich demnach auf der Ebene der Gewissheit bzw. Zuversicht zu bewegen. Wogegen Vertrauen immer nur bestimmten Individuen und deren spezifischen Handeln entgegenbracht werden kann. (Offe (2001a), S. 245)

3. Theorie: Vertrauen

Informationsgrundlage für die Verlässlichkeit dieser Beziehungen nur geringe Qualität aufweist. (Offe (2001a), S. 249ff; S. 252f)

Darüber hinaus „substituiert [Vertrauen; d. Verf.] Ressourcen sozialer Kontrolle und mindert Transaktionskosten“ (Offe (2001a), S. 257). Wer vertraut, für den wird es somit überflüssig ständig zu überprüfen, ob der Kooperationspartner im Sinne des Vertrauenden handelt oder nicht. Die Kontrolle der Loyalität des Kooperationspartners wird durch einen Glauben an seine Redlichkeit und Verlässlichkeit ersetzt. (Offe (2001a), S. 251)

Grundlage der Vertrauensvergabe bleibt die subjektiv wahrgenommene Vertrauenswürdigkeit des Vertrauensnehmers. Diese generiert sich aus Erfahrungen mit dem Vertrauensnehmer, die auf dessen zukünftiges Handeln zu schließen erlauben. Neben früheren Erfahrungen hält OFFE auch die Interaktionsgeschichte von Vertrauensgeber und Vertrauensnehmer für wichtig. Sie gibt Aufschluss darüber, welches der Anlass für die Vergabe von Vertrauen war. Die Kontinuität einer Vertrauensbeziehung ist dann entweder als Ergebnis einer moralischen Reziprozitätsverpflichtung zu verstehen oder als Rezeption eigener Interessen beim Gegenüber (Offe (2001a), S. 253ff). In beiden Fällen tritt das Paradox des Vertrauens zu Tage:

„Die Gewährung von Vertrauen kann genau das Verhalten erzeugen, das logisch gesehen seine Bedingung zu sein scheint.“ (Gambetta (2001), S. 236)

Denn damit Vertrauen gewährt wird, muss eigentlich bereits Vertrauen zwischen den beteiligten Parteien bestehen bzw. eine der beiden Parteien muss der anderen gegenüber in eine riskante Vorleistung treten.

Wie schon weiter oben beschrieben, besteht bei der Vergabe von Vertrauen immer auch das Risiko des Irrtums bzw. des Verlusts. Ein Irrtum kann beispielsweise dann geschehen, wenn entweder der Vertrauensempfänger oder der Vertrauensspender ein rein strategisches Interesse an einer Vertrauensbeziehung haben. Eine der beteiligten Parteien versucht hierbei, sich einen einseitigen Vorteil zu verschaffen, indem sie eine moralische Verpflichtung oder ein geteiltes Interesse nur vorgibt. In solchen Fällen wird das dem Vertrauen zugrundeliegende Reziprozitätsprinzip ausgehebelt. Ein Verlust stellt sich bei einem Vertrauensmissbrauch ein. Hierbei wird der Vertrauensspender vom Vertrauensempfänger betrogen, was einen Verlust des Vertrauens zur Folge hat. In beiden Fällen ist das Ergebnis in der Regel ein Abbruch der Vertrauensbeziehung. (Offe (2001a), S. 255ff)

Die Entscheidung, Vertrauen zu schenken, ist abhängig davon, wie hoch das Risiko eines Verlustes ist bzw. wie groß die Vorteile sind, die sich aus dem Eingang einer Vertrauensbeziehung ergeben. Sind diese beiden Größen als ausreichend hoch zu bewerten, bedarf die Vertrauensvergabe keiner tiefergehenden Kalkulation. Sind die Verhältnisse von Risiken und Vorteilen nicht ohne weiteres eindeutig bestimmbar, so erfolgt eine gründlichere Abwägung. (Offe (2001a), S. 252) Vertrauen ist damit, wie OFFE beschreibt, ein „Optimierungsproblem“ (Offe (2001a), S. 253), das sich bei seiner Lösung nur auf ungesicherte Informationen stützen kann:

„Vertrauen ist eine Überzeugung, die sich auf die Wahrnehmung von Vorstellungen über die Eigenschaften anderer Personen gründet.“ (Offe (2001a), S. 253)

Der Eingang von Vertrauensbeziehungen ergibt sich somit allein auf der Basis einer subjektiven und möglicherweise negativ oder positiv verzerrten Einschätzung anderer Personen.

3. Theorie: Vertrauen

Misstrauen definiert OFFE als „eine das Handeln lenkende Einschätzung einer sozialen Interaktion als riskant“ (Offe (2001a), S. 249). Das Risiko ergibt sich entgegengesetzt zu Vertrauen allerdings nicht aus dem Unvermögen des Vertrauensnehmers im Sinne des Vertrauensspenders zu handeln, sondern aus der Wahrnehmung der mangelnden Bereitschaft des Vertrauensnehmers dies zu tun.

Die Basis von Vertrauen stellen positive Erfahrungen mit anderen Personen dar, die diese als vertrauenswürdig ausweisen. Solche Erfahrungen werden in der Regel durch Interaktion mit konkreten Personen aus dem näheren sozialen Umfeld generiert. Entsprechend ist die einfachste Art Vertrauen zu schenken, indem man den Personen vertraut, denen man schon immer vertraut hat. Oder mit anderen Worten: Man vertraut in der Regel solchen Personen, denen gegenüber schon eine gewisse Vertrautheit besteht. (Offe (2001a), S. 260ff)

Nun besteht die Lebenswelt nicht allein aus vertrauten Personen, sondern vielmehr aus einer Mehrzahl von Personen, die dem einzelnen Individuum vollkommen unbekannt sind. Trotzdem ist gerade in der modernen Welt das Vertrauen zu Fremden von besonderer Wichtigkeit, sind wir doch nicht mehr nur in kleinen autarken Gemeinschaften organisiert, sondern in ein vielschichtiges Gesellschaftssystem integriert, das ohne die Kooperation zwischen sich einander fremden Personen nicht lebensfähig ist. Vertrauen zu Fremden kann dabei nicht auf die althergebrachte Art der fortlaufenden Interaktion miteinander und der daraus resultierenden Erkenntnis über die Vertrauenswürdigkeit des anderen entstehen. Eine Gewährung von Vertrauen auf Basis einer kontinuierlichen Prüfung der Redlichkeit des anderen und dem damit einhergehenden Abbau von Argwohn kann nicht hergestellt werden, da Kontakte mit Fremden meist so flüchtig sind, dass sie nicht zu Bekannten gemacht werden können. Es fehlt die Zeit, längerfristige Beobachtungen anzustellen, die Rückschlüsse auf das potentielle zukünftige Verhalten des Fremden zulassen. Die Flüchtigkeit der Kooperation mit fremden Personen bringt es außerdem mit sich, dass dem Reziprozitätsprinzip des Vertrauens keine Rechnung mehr getragen werden kann. Ein erwiesener Gefallen kann oft nur mit Mühe der Person vergolten werden, die ihn erwiesen hat. (Offe (2001a), S. 260ff) Ein grundlegenderes Problem ergibt sich darüber hinaus aus der Menge der Personen innerhalb einer Gemeinschaft, mit denen das Individuum niemals in Kontakt treten kann bzw. wird und von deren Handeln bzw. deren Entscheidungen es trotzdem abhängig ist. Mit solchen Personen kann weder kooperiert noch kommuniziert werden, was ein unmittelbares Vertrauen unmöglich macht. Das Individuum sieht sich hier in besonderer Weise dem Problem der Kontingenz gegenüber. Es besteht ständig die Gefahr, dass ein anderer seine Freiheit des Handelns benutzt, bewusst oder unbewusst, sodass für ein anderes Individuum ein Nachteil oder Verlust entsteht. (Offe (2000), S. 126ff; Offe (2001a), S. 263; Offe (2001b), S. 171ff)

Offensichtlich besteht innerhalb der Gesellschaft die Notwendigkeit eines verallgemeinerten Vertrauens. Eines Vertrauens, „das großen und unspezifischen Kategorien von ‚fremden‘ Personen entgegengebracht wird“. Solches Vertrauen definiert OFFE als einen „Rückstand, der sich bildet, nachdem Gründe für Misstrauen methodisch widerlegt sind“. (Offe (2001a), S. 262) Diese Kategorie unterscheidet sich deutlich vom weiter oben beschriebenen Vertrauen personalen Charakters. Mit anderen Worten: Die Ausdifferenzierung moderner Gesellschaften und die damit verbundene defizitäre Entwicklung der Möglichkeiten zur personalen Vertrauensbildung macht ein *generalisiertes Vertrauen* notwendig, das mit Verfahren der Mikroebene nicht etabliert werden kann, sondern auf anderem Wege zustande kommen muss. (Offe (2001a), S. 263)

3. Theorie: Vertrauen

„What we need in order to resolve the tension between the desirable and the frightening aspects of popular sovereignty is *trust* – trust, that is, in the reasonableness, informed judgement, and good intentions of our fellow citizens whom we recognize as legitimate co-authors of the law that eventually will bind all of us.“ (Offe (2001b), S. 172f)

Um das Ziel horizontalen Vertrauens zwischen den Mitgliedern einer Gesellschaft zu erreichen, schlägt OFFE drei prinzipielle Möglichkeiten vor: Integration durch (1) populistische Eliten, (2) Entwicklung einer gemeinschaftlichen Identität oder (3) ein System von Institutionen. (Offe (2001a), S. 265ff; Offe (2000), S. 130)

OFFE versteht unter populistischen Eliten demokratisch gewählte Volksvertreter, deren besondere Eigenschaft es ist, „charismatische Eigenschaften zur Schau zu stellen und institutionell vermittelte Popularität und Akklamation hervorrufen“ (Offe (2001a), S. 284). Ihre integrative Kraft entfalten sie durch die Thematisierung nichtpolitischer Inhalte, wie beispielsweise die Einhaltung und Verteidigung gesellschaftlicher Werte und Normen, und durch den Einsatz identitätsbildender Maßnahmen zur Vereinigung unterschiedlicher gesellschaftlicher Gruppen und Strömungen. Gesellschaftliche Integration ist damit allerdings an eine bestimmte Person und deren Funktion gebunden, ein Vertrauen in die Mitbürger stellt sich nach OFFEs Ansicht nicht ein. (Offe (2000), S. 130; Offe (2001a), S. 265, S. 284f)

Unter der Entwicklung einer gemeinschaftlichen Identität soll verstanden werden, „dass die Gesamtheit der Bürger zu Angehörigen einer verpflichtenden und vertrauensbildenden politischen Gemeinschaft erklärt werden“ (Offe (2001a), S. 266). Gesellschaftliche Integration entsteht hierbei auf Basis eines *kategorischen Vertrauens*:

„Dann wird *Zugehörigkeit* zu einer nicht erfahrungsabhängigen Regel, nach der Vertrauenswürdigkeit zuerkannt wird. ‚Ich‘ vertraue ‚dir‘ dann nicht, weil ich dich ‚kenne‘, sondern deduziere deine Vertrauenswürdigkeit aus deiner mir erkennbaren Zugehörigkeit zu einer sozialen Kategorie, deren Angehörige insgesamt für einander als vertrauenswürdig gelten.“ (Offe (2001a), S. 271)

Diese Integrationsmethode wirkt in zwei Richtungen gleichzeitig. Auf der einen Seite ermöglicht sie die Verbindung auch unterschiedlicher gesellschaftlicher Gruppen, gleichzeitig sorgt sie aber im gleichen Maße auch für eine deutliche Abgrenzung von anderen gesellschaftlichen Gruppen, was in einem kategorischen Misstrauen diesen gegenüber gipfeln kann. In einer ausdifferenzierten modernen Gesellschaft mit vielen unterschiedlichen sozialen Gruppen und Strömungen ist eine Integration auf Basis einer gemeinsamen Identität oft nur schwer umzusetzen. (Offe (2001a), S. 266f, S. 271ff)

Institutionen wirken sich mittels der von ihnen vertretenen und aufgestellten institutionalisierten Regeln auf das Vertrauen zu anderen Menschen aus. Grundvoraussetzung für die Generierung *generalisierten Vertrauens* aus Institutionen ist dabei, dass die fraglichen Institutionen „nicht nur die Fähigkeit haben, ‚mich‘ zu verpflichten, sondern (als notwendige Voraussetzung hierfür) in mir das Vertrauen wecken, dass ‚alle anderen‘, obwohl ich diese nicht ‚kenne‘, sich ebenso verpflichtet fühlen werden wie ich“ (Offe (2001a), S. 277). Die vertrauensstiftende Kraft von Institutionen ergibt sich damit aus der Folgebereitschaft eines einzelnen Akteurs gegenüber institutionalisierten Regeln unter Berücksichtigung der Folgebereitschaft von Dritten gegenüber diesen Regeln. Außerdem spielt die subjektiv wahrgenommene Sinnhaftigkeit der Institutionen und die Annahme eines mit anderen Akteuren

3. Theorie: Vertrauen

geteilten institutionellen Umfeldes eine Rolle.¹⁴ (Offe (2000), S. 127f, S. 130; Offe (2001a), S. 265f, S. 274)

Vertrauensstiftende Institutionen stecken einen Verhaltensspielraum ab, an dem sich das Handeln der verschiedenen Akteure orientiert. Das Handeln der anderen kann dadurch antizipiert und mit Sanktionen belegt werden. In der Folge reduziert sich das Risiko eines Vertrauensmissbrauchs:

„Wer das tun würde, so kann ich als Beteiligter einer institutionell geregelten sozialen Interaktion unterstellen, der würde zwar unter Umständen Vorteile erlangen und mich ausbeuten können, zugleich aber diese Vorteile damit bezahlen müssen, dass er sich außerhalb des gegebenen institutionellen Zusammenhanges stellt und sich insofern ‚unmöglich‘ macht, das heißt sich als Objekt von Vertrauensbeziehungen selbst liquidiert. Institutionen leisten eine Art Risikoausgleich: (...) Dieser eingebaute Abschreckungseffekt wird sogar noch verstärkt, wenn Institutionen, wie es zumindest bei formalen Institutionen die Regel ist, nicht nur das Handeln orientieren, sondern auch Abweichungen sanktionieren. Der Schaden für den, der Vertrauen missbraucht, wird größer, und das Erteilen von Vertrauen wird weniger riskant.“ (Offe (2001a), S. 279)

Zusammenfassend stellt OFFE fest, dass sich die Katalysatorfunktion von Institutionen bezüglich *generalisiertem Vertrauen* auf die Vermittlung und Überwachung der Werte *Wahrheit* und *Gerechtigkeit* bezieht. *Generalisiertes Vertrauen* stellt sich hiernach ein, wenn Institutionen einen Verstoß gegen die angesprochenen Werte wirksam sanktionieren können und zusätzlich aktiv an der Verbreitung dieser Werte beteiligt sind. Bezüglich *Wahrheit* bedeutet dies eine Sanktionierung von Täuschung und die Sicherstellung von *Vertragstreue*. Bezüglich *Gerechtigkeit* stehen die Unterbindung von Diskriminierung und die aktive Förderung des Gleichheitsgrundsatzes im Vordergrund. (Offe (2001a), S. 280ff)

„Institutionen generieren Vertrauen unter Fremden (...), wenn sie als Verkörperung dieser Werte wahrgenommen werden und ihnen die Fähigkeit zuerkannt wird, Akteure sowohl im Sinne dieser Werte zu motivieren wie abweichend motiviertes Handeln zu sanktionieren.“ (Offe (2001a), S. 283)

3.2.3.1.3 Generalisierte Reziprozität (Newton)

NEWTON definiert Vertrauen allgemein als „the actor’s belief that, at worst, others will not knowingly or willingly do him harm, and at best, that they will act in his interests“ (Newton (2001), S. 202) bzw. „the belief that others will not deliberately or knowingly do us harm, if they can avoid it, and will look after our interests, if this is possible“ (Delhey/Newton (2005), S. 311).¹⁵ Vertrauen bezieht sich dabei auf eine uns unbekannte Zukunft und beinhaltet deshalb ein Risiko, das auf der Unsicherheit bezüglich des zukünftigen Verhaltens anderer fußt. Zwar bildet Wissen über andere die Grundlage von Vertrauen, sodass je nach Wissensstand entweder vertraut oder misstraut wird, gleichzeitig sind wir hierbei aber immer auch mit dem Induktionsproblem konfrontiert. Unser Wissen über andere ist niemals vollständig, sodass auch unseren Urteilen über andere eine gewisse Unsicherheit innewohnt. Die Etablierung von Vertrauen benötigt

¹⁴ FUHSE und SCHAAL sprechen in diesem Zusammenhang von einer „Kanonisierung gesellschaftlicher Einstellungen und Werte hin zu den institutionellen Leitideen“. Institutionen sorgen in diesem Sinne für „eine gewisse Gleichorientierung individueller Einstellungen“ und eine indirekt daraus resultierende Stärkung des Gemeinschaftsgefühls innerhalb der betroffenen Gesellschaft. (Fuhse/Schaal (2005), S. 57)

¹⁵ NEWTONs Definition politischen Vertrauens lässt sich auf der gleichen Ebene ansiedeln: „For present purpose it is enough to define political trust as the belief that those in authority and with power will not deliberately or willingly do us harm, if they can avoid it, and will look after our interests, if this is possible.“ (Newton (2008), S. 242)

3. Theorie: Vertrauen

deshalb nach NEWTONs Ansicht „a leap of faith“ (Newton 1999a), S. 171). Nur mit Hilfe eines solchen einseitigen Vertrauensvorschlusses lässt sich eine Vertrauensbeziehung initiieren. Mit Vertrauen verbunden ist des Weiteren die Annahme von Reziprozität. Das heißt: Es wird erwartet, dass erwiesenes Vertrauen vergolten wird. NEWTON nimmt hierbei das Prinzip einer generalisierten Reziprozität an, welches davon ausgeht, dass ein Gefallen nicht unbedingt von der Person erwidert werden muss, der er erwiesen wurde, sondern dafür prinzipiell jede Person in Frage kommt. Generalisierte Reziprozität geht somit davon aus, dass sich für einen Gefallen irgendwann in der Zukunft von irgendjemandem revanchiert wird. (Newton (1999a), S. 171; Newton (1999b), S. 4)

NEWTON differenziert Vertrauen in *thick trust* und *thin trust*¹⁶, wobei ersteres *partikularem* bzw. *spezifischem* Vertrauen und letzteres *generalisiertem* bzw. *diffusem* Vertrauen entsprechen. *Thick trust* entsteht in engen Gemeinschaften und ergibt sich aus intensivem und täglichem Kontakt mit Personen des gleichen Stammes, der gleichen Klasse oder der gleichen ethnischen Gruppe. Kennzeichen solcher Gemeinschaften ist ihre Homogenität. Sichergestellt wird *thick trust* durch Methoden sozialer Kontrolle und soziale Sanktionen. *Thin trust* dagegen ergibt sich auf Basis loser und unregelmäßiger Kontakte und basiert weniger stark auf persönlicher Bekanntheit. (Delhey/Newton/Welzel (2011), S. 786; Newton (1999b), S. 14ff; Newton (2008), S. 246)

Während die von NEWTON als *communal society* bezeichnete Vormoderne noch Strukturen aufwies, die von *thick trust* geprägt waren, verliert solches Vertrauen in modernen Gesellschaften zusehends an Bedeutung und wird durch *thin trust* ersetzt. Innerhalb moderner Gesellschaften spielen sich Interaktionen nicht mehr nur in kleinen Gemeinden ab, in denen man sich wechselseitig bekannt ist, vielmehr basiert der Großteil von Interaktionen auf Kontakt zu Fremden. Im Zuge von erweiterter geographischer Mobilität und fortschreitender Individualisierung gestaltet sich die moderne Gesellschaft heterogener und hat sich auch im Bereich der Aufgabenteilung stark ausdifferenziert. Gesellschaftsrelevante Pflichten und Rollen werden nun nicht mehr komplett vom direkten Umfeld eines Individuums wahrgenommen, sondern verteilen sich auf eine Vielzahl von Personen, die dem Individuum möglicherweise vollkommen unbekannt sind. Um ein Funktionieren ausdifferenzierter moderner Gesellschaften zu ermöglichen, ist es daher unumgänglich auch Fremden Vertrauen entgegenzubringen. (Delhey/Newton (2005), S. 311f; Newton (1999a), S. 174; Newton (2008), S. 246)

Als mögliche Ausgangspunkte solchen *thin trust* macht NEWTON im Wesentlichen zwei Quellen aus: (1) die Mitgliedschaft in formellen und informellen Netzwerken und (2) den Einfluss von Institutionen auf das tägliche Miteinander. Netzwerke werden dabei durch sogenannte interne und externe Effekte wirksam. Als interner Effekt wird die Sozialisation der Mitglieder im Sinne einer demokratischen Kultur bezeichnet. Innerhalb des Netzwerkes erlernen die Mitglieder die „civic virtues of trust“ (Newton (1999b), S. 11), die sie zu Reziprozität, Kooperation, Kompromissfähigkeit und Empathie erziehen. Die Mitglieder entwickeln darüber hinaus ein Gefühl für Gemeinschaftsinteressen und Gemeinschaftsgüter. Externe Effekte ergeben sich aus dem Aufeinandertreffen unterschiedlicher sozialer Gruppen innerhalb der Netzwerke. Auf diese Weise werden sowohl Zusammenhalt als auch Wettbewerb zwischen den Gruppen und ihren Interessen gefördert. Zusätzlich wird der einzelne Bürger über das Netzwerk mit dem politischen System und seinen Entscheidungsorganen verbunden. Vertrauen entwickelt

¹⁶ NEWTON nimmt hierbei Bezug auf die Konzeption *mechanischer* und *organischer Solidarität* von DURKHEIM sowie die Konzeption sogenannter *strong* und *weak ties* von GRANOVETTER. (Dürkheim (1992), S. 118ff; Granovetter (1973))

3. Theorie: Vertrauen

sich hier demnach aus sozialer Solidarität. (Delhey/Newton (2002), S. 6ff; Newton (1999a), S. 171ff; Newton (1999b), S. 11, S. 15; Newton (2001), S. 206f; Newton/Norris (2000), S. 66) Gleichwohl konstatiert NEWTON, dass nicht eindeutig festgelegt werden kann, ob der Kausalpfeil zwischen *generalisiertem Vertrauen* und Netzwerkmitgliedschaft in Richtung des Vertrauens oder der Mitgliedschaft weist. Er nimmt allerdings an, dass Letzteres gilt:

„It seems at least as likely that trusting people tend to join organizations as the other way around. The relationship is likely to be reciprocal, but strongest path probably runs from trusting to joining associations (...). The argument for this supposition is as follows: we know that social winners tend to be trusting, and that joiners of voluntary organizations are often winners with comparatively high income, education, and social class. It seems more likely that the background and experience of being a social winner encourages both joining and trusting. It is less plausible to argue that people are trusting because they have learned this attitude in their voluntary organizations, although membership may reinforce pre-existing levels of trust.“ (Newton (2001), S. 207)

Zusätzlich führt er an, dass es zwar eine Vielzahl von Vereinigungen gibt, die durchaus eine entsprechende Wirkung auf die Verteilung und Entstehung generalisierten Vertrauens haben könnten, aber nur vergleichsweise wenige Menschen in solchen Verbänden organisiert sind, als dass ein besonderer Einfluss ableitbar wäre. Vielmehr verhält es sich sogar oft so, dass es sich lediglich um eine passive Mitgliedschaft handelt, bei der Mitgliedsbeiträge aus symbolischen Gründen oder wegen eines besonderen Nutzens gezahlt werden. (Newton (1999a), S. 172; Newton (1999b), S. 6, S. 8, S. 12)

Institutionen wirken sich in zweierlei Hinsicht auf Vertrauen aus, letztlich sind aber beide auf die Sicherung sozialer Ordnung rückführbar. Institutionen stellen in erster Linie Kontrollorgane dar und sichern auf diese Weise die Vertrauenswürdigkeit von Bürgern, aber auch politischen Entscheidungsträgern. Bewerkstelligt wird dies mit Hilfe der von den Institutionen bereitgestellten juristischen und bürokratischen Mechanismen. Auf einer zweiten Ebene garantieren Institutionen aber auch jedem Bürger die gleichen Rechte und Pflichten. Die Folge ist ein gesteigertes Vertrauen in die Mitbürger, das dieser Gleichheit entspringt. Entsprechend des sogenannten *rainmaker effects* wirken sich Institutionen dabei auf alle Mitglieder einer Gesellschaft gleichermaßen und in gleicher Weise aus. (Newton (2008), S. 255f; Newton/Norris (2000), S. 53; Newton/Zmerli (2011), S. 169)

„In the same way that the rain from heaven falls on the just and the unjust alike, so also the operations of social institutions affect all citizens to greater or lesser extent, irrespective of whether they are trusting individuals or not. Behaviour towards fellow citizens is likely to be more honest and trustworthy where there is an impartial police force, a just legal system, and an honest bureaucracy. There is usually a strong correlation between social trust and confidence in the police, courts, and state bureaucracy because these are precisely the public institutions that are supposed to maintain an impersonal, universalistic, and rule-bound social order that encourages trustworthy behavior among citizens.“ (Newton (2008), S. 255)

Zuversicht gegenüber Institutionen¹⁷, insbesondere solchen, die soziale Ordnung garantieren, und *generalisiertes Vertrauen* sind eng miteinander verbunden. Allerdings stellt sich auch hierbei wieder die Frage der Kausalität: Führen Institutionen zu einem Anstieg bzw. der

¹⁷ Nach Auffassung NEWTONs sollte in Bezug auf Institutionen statt von *Vertrauen (trust)* von *Zuversicht (confidence)* gesprochen werden. Während der Begriff des Vertrauens für Einstellungen gegenüber Individuen reserviert ist und sich damit im Bereich persönlicher Beziehungen bewegt, bezieht sich der Begriff der Zuversicht auf Gesellschaft und den Bereich des Politischen. Vertrauen generiert sich dabei aus persönlicher Erfahrung und persönlicher Bekanntschaft, während sich Zuversicht vornehmlich auf Informationen aus zweiter Hand stützt. (Newton (1999a), S. 179; Zmerli/Newton/Montero (2007), S. 39f)

Konstitution *generalisierten Vertrauens* oder hat im Gegenteil eine Erhöhung des *generalisierten Vertrauens* einen stabilisierenden Effekt auf Institutionen? Gesichert scheint, dass Zuversicht gegenüber Institutionen eine wichtige Rolle für die politische Stabilität und Integration in modernen Gesellschaften spielt. (Newton/Norris (2000), S. 53; S. 66f, S. 71f)

3.2.3.2 Institutional-structural Approach

Institutionell-strukturelle Ansätze legen den Fokus auf die Rolle des Staates als Quelle *generalisierten Vertrauens*. So tragen Institutionen zur Produktion *generalisierten Vertrauens* bei, indem sie anomisches Verhalten sanktionieren sowie über die Verbreitung von kulturellen Prinzipien und Normen die Werte und Überzeugungen der Menschen gestalten. Von besonderer Relevanz ist in diesem Zusammenhang die Unterscheidung von sogenannten Repräsentations- und Implementationsinstitutionen, da vornehmlich letztgenannte für die Umsetzung und Durchsetzung normativer Standards zuständig sind. Entsprechend können Judikative und Administration durch ihren oft unmittelbaren Kontakt mit dem Bürger großen Einfluss auf *generalisiertes Vertrauen* ausüben. Voraussetzung für den Beitrag dieser Klasse von Institutionen zu *generalisiertem Vertrauen* ist die Wahrnehmung des Staates als vertrauenswürdig. (Rothstein/Stolle (2002), S. 10ff; Rothstein/Stolle (2007), S. 9ff; Rothstein/Stolle (2008a), S. 281ff; Rothstein/Stolle (2008b), S. 444ff)

3.2.3.2.1 Vertrauensnormen als Folge von „street-level bureaucracy“ (Rothstein/ Stolle)

Nach Auffassung von ROTHSTEIN/STOLLE stellt *generalisiertes Vertrauen* die Kerninstanz sozialen Kapitals dar und ist wie folgt zu charakterisieren:

„Generalized trust indicates the potential readiness of citizens to cooperate with each other and the abstract preparedness to engage in civic endeavors with each other. Attitudes of generalized trust extend beyond the boundaries of face-to-face interactions and incorporate people who are not personally known. These attitudes of trust are generalized when they go beyond specific personal settings in which the partner to be cooperated with is already known. They even go beyond the boundaries of kinship and friendship, and the boundaries of acquaintance.“ (Rothstein/Stolle (2002), S. 2f)

Hauptfunktion *generalisierten Vertrauens* ist es dabei, die Unsicherheit bezüglich einer unbekannten Zukunft abzubauen sowie das Problem der doppelten Kontingenz zu entschärfen. *Generalisiertes Vertrauen* fungiert dann als Moderator, der risikobehaftete soziale Interaktion möglich macht. (Rothstein/Stolle (2002), S. 3) Insbesondere in sogenannten *one-shot-situations*, dem einmaligen Zusammentreffen mit einer Person, oder Situationen, in denen man mit einer großen Zahl von Personen interagieren muss, spielt *generalisiertes Vertrauen* eine große Rolle. Insgesamt trägt es dazu bei, dass Personen eher bereit sind zu kooperieren, was einen gesamtgesellschaftlichen Nutzen nach sich zieht. (Rothstein/Stolle (2007), S. 2)

Zur Erklärung der Genese *generalisierten Vertrauens* existieren nach ROTHSTEIN/STOLLE zwei verschiedene Erklärungsansätze: *society-centered accounts* und *institutional-centered accounts*.

Folgt man den *society-centered accounts*, so ergibt sich *generalisiertes Vertrauen* aus den historischen und kulturellen Erfahrungen eines gesellschaftlichen Gefüges. Das Vertrauen entsteht dabei in der Folge eines *bottom-up*-Prozesses aus der sozialen Interaktion der Bürger,

3. Theorie: Vertrauen

wobei vornehmlich formelle und informelle Vereinigungen und Netzwerke als Katalysator der Vertrauensproduktion dienen. So wird angenommen, dass den Mitgliedern solcher Organisationen durch Gruppenaktivitäten demokratische und kooperative Werte und Normen vermittelt werden, die sich dann durch die Interaktion mit Nicht-Mitgliedern auf die gesamte Gesellschaft übertragen. (Rothstein/Stolle (2002), S. 4f; Rothstein/Stolle (2008a), S. 275; Rothstein/Stolle (2008b), S. 442)

ROTHSTEIN/STOLLE sehen diesen Erklärungsansatz mit gleich vier Problemen konfrontiert (Rothstein/Stolle (2002), S. 5ff; Rothstein/Stolle (2008a), S. 276ff; Rothstein/Stolle (2008b), S. 442f):

- (1) Der erste Kritikpunkt betrifft den Zweck der Netzwerke. Formelle oder informelle Vereinigungen müssen nicht immer das Gemeinwohl im Blick haben. Vielmehr liegt es im Interesse vieler Organisationen sich von anderen Gruppen abzugrenzen und dabei Misstrauen gegenüber diesen Gruppen zu schüren. Es findet somit eine Exklusion von Nicht-Mitgliedern statt, die sich in einem starken *ingroup*-Vertrauen und einem starken *outgroup*-Misstrauen manifestiert. Es muss daher zwischen sogenannten *bridging*- und *bonding*-Organisationen unterschieden werden. Erstere ermöglichen und fördern Kontakte zu Personen, die außerhalb des Netzwerkes stehen und sich mitunter deutlich von den Mitgliedern unterscheiden, während letztere lediglich Kontakte zu Personen herstellen, die den Mitgliedern des Netzwerkes entsprechen. Ein Beispiel für die Exklusionswirkung und die Abkehr vom Gemeinwohl sind die Annehmlichkeiten und Vorteile die Organisationsmitglieder gegenüber Nicht-Mitgliedern genießen, die oft aus gesamtgesellschaftlichen Ressourcen abgezogen und somit auf Kosten der verbleibenden Gesellschaft gewährt werden.
- (2) Der zweite Kritikpunkt schließt an den ersten an und betrifft die Details der Übertragung des innerhalb der Vereinigung generierten Vertrauens auf die Gesellschaft. Da es sich bei *generalisiertem Vertrauen* um ein Kollektivgut handelt, das zwischen Menschen existiert, d.h. an Interaktion bzw. den Zusammenschluss von Menschen gebunden ist, bleibt unklar, wie Personen außerhalb solcher Kollektive vom dort produzierten Vertrauen profitieren können.
- (3) Der dritte Kritikpunkt betrifft die kausale Richtung der Genese von *generalisiertem Vertrauen*. So sind Mitglieder sozialer Netzwerke in erhöhtem Maße vertrauens- und kooperationsbereit, gleichzeitig ist aber die Wahrscheinlichkeit als sogenannter *truster* einem sozialen Netzwerk beizutreten ungleich höher, als dies als sogenannter *distruster* zu tun. Es handelt sich somit bei der Häufung von Vertrauen innerhalb einer solchen Organisation nicht um einen Effekt der Vertrauensproduktion als Folge der Aktivitäten der Organisation, sondern vielmehr um einen Effekt der Selbstselektion ihrer Mitglieder. Soziale Netzwerke tragen nur vermeintlich zur Vertrauensproduktion bei. In Wirklichkeit haben die Mitglieder schon vor Eintritt in das Netzwerk einen hohen Vertrauenslevel. Hätten soziale Vereinigungen den postulierten Einfluss auf die gesamtgesellschaftliche Produktion von *generalisiertem Vertrauen*, so müssten Regionen mit engen sozialen Netzwerken einen erhöhten Einfluss auf die Generalisierung von Vertrauen haben und gleichzeitig müssten Individuen, die einem solchen Netzwerk beitreten, lernen zu vertrauen. Allerdings zeigen empirische Untersuchungen, dass „a causal flow from joining to trusting is nowhere to be found“ (Rothstein/Stolle (2008a), S. 277).
- (4) Der vierte Kritikpunkt betrifft den historischen Determinismus, der den *society-centered accounts* innewohnt. Die Genese von *generalisiertem Vertrauen* wird dabei als Folge eines

3. Theorie: Vertrauen

historischen Prozesses beschrieben, dessen Anfangspunkt teilweise weit in der Vergangenheit zu verorten ist. Wenn allerdings das Ausmaß *generalisierten Vertrauens* historisch determiniert ist und auf in der Zeit mitunter weitzurückliegenden sozialen Strukturen basiert, so kann es in der Gegenwart nur wenige Optionen zur Manipulation der Produktion von *generalisiertem Vertrauen* geben. Tatsächlich gestaltet es sich so, dass die Zerstörung *generalisierten Vertrauens* eher zu bewerkstelligen ist als dessen Förderung.

Abschließend kommen ROTHSTEIN/STOLLE bezüglich der *society-centered accounts* zu folgendem Fazit:

„The net conclusion from empirical research is that associativeness and social networks may very well be a good thing for many reasons, but they do not seem to produce interpersonal trust and wider norms of reciprocity that benefit the whole society as social capital theory originally implied.“ (Rothstein/Stolle (2008a), S. 278)

Institution-centered accounts führen die Genese *generalisierten Vertrauens* auf die Existenz politischer und judikativer Institutionen zurück. So wird angenommen, dass solche Institutionen *generalisiertes Vertrauen*, durch die im Rahmen ihrer Kompetenzen liegenden Möglichkeiten des Eingriffs in die Gesellschaft, sowohl beeinflussen als auch produzieren können. Innerhalb der *institution-centered accounts* unterscheiden ROTHSTEIN/STOLLE zwei Erklärungsmodelle: den *attitudinal approach* und den *institutional-structural approach*. (Rothstein/Stolle (2002), S. 7f; Rothstein/Stolle (2008a), S. 278f; Rothstein/Stolle (2008b), S. 443)

Der *attitudinal approach* untersucht, in welcher Beziehung Einstellungen der Zuversicht in politische Institutionen und Einstellungen des *generalisierten Vertrauens* zueinanderstehen. Hinsichtlich der kausalen Richtung des Zusammenhangs wird dabei häufig angenommen, dass *generalisiertes Vertrauen* einen positiven Effekt auf Zuversicht gegenüber Institutionen hat. So wird vermutet, dass *generalisiertes Vertrauen* zur Etablierung effektiver politischer Institutionen beiträgt, die wiederum eine wirksame Regierungsarbeit ermöglichen, was sich positiv auf das Ausmaß der Zuversicht gegenüber den Institutionen auswirkt. Andere Ansätze gehen hingegen davon aus, dass der Kausalpfeil in die entgegengesetzte Richtung weist und Zuversicht in Institutionen den Prädiktor für *generalisiertes Vertrauen* darstellt. (Rothstein/Stolle (2002), S. 8f; Rothstein/Stolle (2007), S. 8; Rothstein/Stolle (2008a), S. 279f; Rothstein/Stolle (2008b), S. 443)

Nicht zuletzt wegen der Unsicherheit hinsichtlich der kausalen Richtung des Zusammenhangs von *generalisiertem Vertrauen* und Zuversicht in Institutionen sehen ROTHSTEIN/STOLLE den *attitudinal approach* mit Problemen behaftet. Ein weiterer Kritikpunkt ist, dass sich der prognostizierte Zusammenhang zwischen Institutionen und *generalisiertem Vertrauen* nur als Folge einer Wechselbeziehung von Einstellungen manifestiert. Die tatsächlichen Charakteristika der involvierten Institutionen spielen hierbei, wenn überhaupt, nur indirekt eine Rolle. Hieran anschließend wird außerdem kritisiert, dass die bisherige Forschung Vertrauen gegenüber Institutionen relativ undifferenziert betrachtet hat. So werden verschiedene Formen des Institutionenvertrauens unter einen Oberbegriff subsumiert. Auf diese Weise gehen auch Institutionen in die Analyse mit ein, die wenig bis gar nichts zu *generalisiertem Vertrauen* beitragen und damit den Zusammenhang zwischen Institutionenvertrauen und *generalisiertem Vertrauen* unterminieren. (Rothstein/Stolle (2001), S. 9; Rothstein/Stolle (2002), S. 9; Rothstein/Stolle (2007), S. 9; Rothstein/Stolle (2008a), S. 280; Rothstein/Stolle (2008b), S. 448f)

Der von ROTHSTEIN/STOLLE vorgeschlagene *institutional-structural approach* stellt den Beitrag des Staates zur Generierung *generalisierten Vertrauens* in den Mittelpunkt.

3. Theorie: Vertrauen

Grundvoraussetzung hierfür ist, dass der Staat von den Bürgern als vertrauenswürdig angesehen wird. Die Befähigung zur Beeinflussung des Levels *generalisierten Vertrauens* ergibt sich dabei vornehmlich aus der Kompetenz des Staates, Vertragssicherheit zu gewähren, Rechte und Regeln zur Sanktionierung devianten Verhaltens durchzusetzen sowie die Integration und Partizipation der Bürger zu unterstützen. (Rothstein/Stolle (2002), S. 10; Rothstein/Stolle (2007), S. 10; Rothstein/Stolle (2008a), S. 281)

Ausgehend von der Kritik am *attitudinal approach* unterscheiden ROTHSTEIN/STOLLE hierbei zunächst zwischen sogenannten Repräsentations- und Implementationsinstitutionen. Repräsentationsinstitutionen tragen ihrer Auffassung nach kaum zur Generierung *generalisierten Vertrauens* bei. Solche Institutionen sind in der Regel durch eine parteispezifische Voreingenommenheit charakterisiert, dabei besteht ihre Hauptaufgabe aus der Implementation der Parteiideologie der jeweiligen Regierungspartei. Die Folge ist, dass vornehmlich die Anhänger dieser Parteiideologie Vertrauen entwickeln, während die Opposition eher zu Misstrauen neigt. Zudem ist das Vertrauen in Repräsentationsinstitutionen oft nur temporärer Natur und zielt auf kurzfristige und spezifische politische bzw. ökonomische Projekte, wobei es an die Person bestimmter Amtsträger geknüpft ist. Entsprechend ist die Verbindung zwischen *generalisiertem Vertrauen* und Zuversicht gegenüber dieser Art von Institutionen in der Regel schwach. (Rothstein/Stolle (2001), S. 7; Rothstein/Stolle (2002), S. 11f; Rothstein/Stolle (2003), S. 193f; Rothstein/Stolle (2007), S. 10f; Rothstein/Stolle (2008a), S. 282f; Rothstein/Stolle (2008b), S. 444)

Eine weit größere Rolle für die Generierung und Aufrechterhaltung von *generalisiertem Vertrauen* spielen nach ROTHSTEIN/STOLLE die sogenannten Implementationsinstitutionen. Hierzu gehören insbesondere die Judikative und die Administration, deren Kernaufgaben die Ermittlung und Sanktionierung devianten bzw. anomischen Verhaltens sowie die Umsetzung staatlicher Politikvorgaben darstellen. Der Hauptunterschied zu Repräsentationsinstitutionen besteht allerdings in ihrer Dauerhaftigkeit und Stabilität über die Zeit. Anders als bei Repräsentationsinstitutionen besitzen Implementationsinstitutionen eine größere Beständigkeit, so ist die personelle Besetzung von Ämtern innerhalb dieser Institutionen oft unabhängig von Wechseln, die durch den Ablauf der Legislaturperiode induziert sind. Diese Beständigkeit erlaubt es solchen Institutionen einen erheblichen Einfluss auf die Sozialisation der Bürger zu nehmen. Hinzu kommt außerdem, dass Judikative und Administration eine größere Bedeutung für das gesellschaftliche Tagesgeschäft besitzen als andere Arten von Institutionen. Der Kontakt zwischen Bürger und Institution verläuft hierbei viel direkter, da Bürger unmittelbar mit Vertretern der jeweiligen Institutionen interagieren können.¹⁸ Die Einflussnahme der Institutionen auf *generalisiertes Vertrauen* verläuft dabei entlang der Repräsentation grundlegender gesellschaftlicher Ideale. Institutionen sollten hiernach bei der Interaktion mit der Gesellschaft die Prinzipien der Unparteilichkeit, Objektivität, Fairness, Neutralität und Effizienz befolgen. (Rothstein/Stolle (2001), S. 7; Rothstein/Stolle (2002), S. 11ff; Rothstein/Stolle (2003), S. 194; Rothstein/Stolle (2007), S. 11; Rothstein/Stolle (2008a), S. 283f)

Die Befolgung dieser Prinzipien wird hinsichtlich der Generierung und Aufrechterhaltung *generalisierten Vertrauens* auf zwei Arten und Weisen wirksam (Rothstein/Stolle (2001), S. 7ff, S. 38; Rothstein/Stolle (2002), S. 11ff, S. 29; Rothstein/Stolle (2003), S. 194f; Rothstein/Stolle (2007), S. 11ff, S. 29; Rothstein/Stolle (2008a), S. 283f):

¹⁸ STOLLE/ROTHSTEIN sprechen in diesem Zusammenhang von „*street-level bureaucracy*“. Als Beispiele werden das Rechtssystem und das Wohlfahrtssystem genannt. (Rothstein/Stolle (2001), S. 12ff; Rothstein/Stolle (2002), S. 12; Rothstein/Stolle (2003), S. 195ff; Rothstein/Stolle (2007), S. 15; Rothstein/Stolle (2008a), S. 283)

3. Theorie: Vertrauen

- (1) Zunächst dient die Einhaltung der obengenannten Maximen zur Aufrechterhaltung der Zuversicht gegenüber den Institutionen selbst. Richtet sich die Arbeitsweise der Institutionen anhand von Unparteilichkeit, Objektivität, Fairness und Neutralität aus, so kann der Bürger eine unvoreingenommene Behandlung durch die Institutionen erwarten. Entsprechend kann der Bürger davon ausgehen, dass seine Interessen nicht gegenüber den Interessen anderer Personen benachteiligt oder bevorteilt werden. Der Grundsatz der Effizienz bezieht sich dagegen auf die Sanktionierung von Verletzungen der gesellschaftlichen Regeln und Normen. So soll abweichendes oder opportunistisches Verhalten von den entsprechenden Institutionen erfolgreich geahndet werden, um den Bürgern ein Gefühl von Schutz und Sicherheit zu vermitteln. Die Verbindung aus der Befolgung des Gleichheitsgrundsatzes sowie der Gewährleistung von persönlicher Sicherheit führt dann zu einer Wahrnehmung der Institutionen als vertrauenswürdig.
- (2) Daneben werden die Normen und Prinzipien, nach denen die Institutionen arbeiten, auf die gesamte Gesellschaft übertragen. Diese Generalisierung ergibt sich dabei aus der Annahme, dass wenn sich staatliche Institutionen an bestimmte Verhaltensprinzipien halten, dies auch für die meisten Menschen innerhalb der jeweiligen Gesellschaft angenommen werden und diesen somit vertraut werden kann. Entsprechend gilt aber auch, dass wenn sich die Institutionen nicht an diese Grundregeln halten, die Schlussfolgerung naheliegt, dass auch die Mitbürger sich nicht an diese Regeln halten werden und somit vertrauensunwürdig sind. Hierbei kommt den sogenannten *street-level bureaucrats* eine besondere Rolle zu. Diese stellen die unmittelbare Schnittstelle zwischen Institution und Bürger dar, sodass deren Verhalten als repräsentativ für die dahinterstehende Institution angenommen wird. Korruptes Verhalten seitens dieser Beamten hat dann negative Auswirkungen auf das *generalisierte Vertrauen*. Sie werden quasi als Gradmesser für die generellen moralischen Standards einer Gesellschaft angenommen. Die Wahrnehmung von Justiz und Administration als objektiv, fair und effizient führt darüber hinaus dazu, dass es als weniger wahrscheinlich angesehen wird, dass andere Bürger versuchen werden, sich mittels illegaler Methoden Vorteile zu verschaffen. Es herrscht vielmehr der Glaube vor, dass man mit betrügerischem und korruptem Verhalten nicht durchkommt und deshalb davon abgesehen wird. Gleichzeitig führen als unfair oder korrupt wahrgenommene Justizbehörden oder Verwaltungsorgane zu einer Klassifizierung des gesellschaftlichen Systems als vertrauensunwürdig. Ein solches als vertrauensunwürdig eingestuftes System macht es in letzter Instanz für den einzelnen Bürger notwendig, sich ebenfalls vertrauensunwürdig zu verhalten, da er annehmen muss, dass auch seinen Mitbürger nicht vertraut werden kann.

Der zwischen Institutionen und *generalisiertem Vertrauen* wirksam werdende kausale Mechanismus lässt sich nun anhand von vier Gesichtspunkten charakterisieren (Rothstein/Stolle (2001), S. 16f; Rothstein/Stolle (2002), S. 18f; Rothstein/Stolle (2003), S. 199f; Rothstein/Stolle (2007), S. 16; Rothstein/Stolle (2008b), S. 446):

- (1) Faire und effektive Institutionen wirken sich positiv auf die Rezeption individueller Sicherheit aus. Besteht ein institutionalisierter Schutz vor betrügerischem Verhalten anderer, so kann den Mitbürgern eher vertraut werden.
- (2) Vom Verhalten der Repräsentanten effektiver und fairer Institutionen lassen sich Rückschlüsse auf das Verhalten der verbleibenden Mitglieder der Gesellschaft ziehen. Gelten diejenigen als vertrauensunwürdig, die für soziale Kontrolle und soziale Ordnung verantwortlich sind, so haben auch die verbleibenden Mitbürger keinerlei Anreiz, sich an

übergeordneten Werten und Normen zu orientieren. Die Folge ist, dass den Mitbürgern ebenso wie den Institutionenrepräsentanten nicht vertraut werden kann.

- (3) Die von fairen und effektiven Institutionen vorgelebten Normen und Werte bilden die Blaupause anhand der das Verhalten der Mitbürger bewertet wird. In einem korrupten System wird daher betrügerisches Verhalten als normkonformes Verhalten angesehen und als Leitlinie für das eigene Verhalten adaptiert. Die Konsequenz ist eine Gesellschaft, in der der Einzelne nach Möglichkeit rein eigennützige Vorteile aus der Interaktion mit anderen oder dem System ziehen möchte. In Summe wird dadurch die Etablierung *generalisierten Vertrauens* verhindert.
- (4) Auch über den direkten Kontakt mit fairen und objektiven Institutionen können positive Effekte auf das *generalisierte Vertrauen* generalisiert werden. Institutionen, die sich an das Prinzip der Unparteilichkeit halten, lassen den Bürger erfahren, dass er bzw. seine Interessen nicht diskriminiert werden. Insbesondere der von ROTHSTEIN/STOLLE als *procedural fairness* (Rothstein/Stolle (2007), S. 11f) bezeichnete Umstand der Gleichheit vor dem Gesetz spielt eine bedeutende Rolle für die Generierung und Aufrechterhaltung *generalisierten Vertrauens*.

Abschließend lässt sich folgende grundlegende Überlegung hinter der von ROTHSTEIN/STOLLE vorgenommenen Unterscheidung zwischen Repräsentations- und Implementationsinstitutionen und deren Einfluss auf *generalisiertes Vertrauen* konstatieren:

„In sum, we argue that there are at least two dimensions along which citizens might judge political institutions: on the one hand they expect representatives of political institutions to function as their agents; on the other hand, they judge policy institutions according to their neutrality, fairness, and impartiality. Moreover, citizens expect more agency and more political bias from political institutions with elected offices, whereas they expect impartiality and an unbiased approach from order institutions. Our claim is that the lack of impartiality of public policy institutions damages generalized trust; and alternatively, institutions' perceived impartiality should support generalized trust.“ (Rothstein/Stolle (2008a), S. 284)

3.2.3.2 Institutionelle Regeln als Katalysator des Vertrauens (Freitag/Bühlmann)

Für FREITAG/BÜHLMANN spielt Vertrauen insbesondere in modernen Gesellschaften eine große Rolle. So sorgt die erhöhte Mobilität und Ausdifferenzierung auf nahezu sämtlichen gesellschaftlichen Ebenen dafür, dass vermehrt mit Personen kooperiert werden muss, die einem nicht persönlich bekannt sind. Vertrauen dient hierbei als Katalysator zur Etablierung wechselseitig vorteilhafter Kooperationsbeziehungen und überbrückt die Unsicherheit hinsichtlich der Absichten des Kooperationspartners. In diesem Sinne ist innerhalb einer modernen Gesellschaft jedwede Kooperation im ökonomischen, politischen oder sozialen Bereich auf ein Mindestmaß an Vertrauen angewiesen. (Freitag/Bühlmann (2005), S. 575ff; Freitag/Bühlmann (2009), S. 1537ff)

In Anlehnung an OFFE definieren FREITAG/BÜHLMANN Vertrauen allgemein als „Überzeugung vom zuverlässigen Handeln eines anderen oder einer Gruppe anderer“ (Freitag/Bühlmann (2005), S. 579) bzw. „as the conviction that the actions of another party (an individual or group) can be depended on“ (Freitag/Bühlmann (2009), S. 1540). Entlang des Bezugsrahmens bzw. Radius des Vertrauens unterscheiden FREITAG/BÜHLMANN außerdem zwei Vertrauensformen, das Nachbereichsvertrauen (*particularistic trust*) und das *generalisierte Vertrauen* (*generalized*

3. Theorie: Vertrauen

trust). Voraussetzung für Nahbereichsvertrauen ist eine persönliche Beziehung zu der Person oder Gruppe, der vertraut werden soll. Es basiert auf gemeinsamen Erfahrungen und daraus abgeleiteten Erwartungen hinsichtlich zukünftiger Austauschprozesse. Solches Vertrauen bezieht sich ausschließlich auf einen engen Kreis vertrauter Personen und orientiert sich an der der jeweiligen Beziehung innewohnenden Reziprozitätswahrscheinlichkeit. (Freitag/Bühlmann (2005), S. 579; Freitag/Bühlmann (2009), S. 1540)

Insbesondere in der modernen Gesellschaft, innerhalb derer man gezwungen ist, mit einer Vielzahl fremder Menschen zu interagieren und zu kooperieren, wobei jeweils ein gewisses Mindestmaß an Vertrauen vorausgesetzt werden muss, kann Vertrauen nicht auf dem Wege wiederholten Kontaktes aufgebaut werden. Es muss demnach noch eine zweite Vertrauensform existieren, die die Interaktion und Kooperation mit Fremden ermöglicht. Diese von FREITAG/BÜHLMANN als *generalisiertes Vertrauen* bezeichnete Form des Vertrauens bezieht sich auf Fremde und unbekannte Gruppen und orientiert sich an altruistischen Verhaltensstandards und dem Gemeinwohl. Dabei wird keinerlei direkte Reziprozität vorausgesetzt, vielmehr kann eine Erwidierung der Vertrauensleistungen auch von anderen Personen zu anderen Zeitpunkten und an anderen Orten geleistet werden. (Freitag/Bühlmann (2005), S. 577, S. 579f; Freitag/Bühlmann (2009), S. 1540)

Generalisiertes Vertrauen kann durch verschiedene Mechanismen generiert werden. Besondere Prominenz besitzen dabei Erklärungen, die sich auf die Teilhabe an gesellschaftlichen Vereinigungen oder die Funktion politischer Institutionen stützen. FREITAG/BÜHLMANN gehen indes davon aus, dass insbesondere politische Institutionen zur Vertrauensentwicklung beitragen, wenn sie von allen Individuen gleichermaßen als universalistisch, machtteilend-konsensual, unbestechlich, unparteiisch und als nicht-kooperatives-Verhalten-sanktionierend angenommen werden. (Freitag/Bühlmann (2005), S. 577; Freitag/Bühlmann (2009), S. 1539) Institutionelle Regeln und Konfliktlösemechanismen fungieren hier als Katalysator für *generalisiertes Vertrauen*. Sie bilden eine Art überindividuellen Kontext, innerhalb dessen jegliche Handlungen und Einstellungen stattfinden bzw. ermöglicht, aber auch sanktioniert werden. Vertrauensbildung auf Basis institutioneller Rahmenbedingungen funktioniert, indem Institutionen Stimuli für gemeinschaftsorientiertes Handeln schaffen und gleichzeitig die Überzeugung nähren, dass auch alle anderen Mitglieder der Gesellschaft, insbesondere solche, die dem Einzelnen unbekannt sind, den gleichen Bestimmungen unterliegen bzw. den gleichen Handlungsmaximen folgen. (Freitag/Bühlmann (2005), S. 581f; Freitag/Bühlmann (2009), S. 1543) FREITAG/BÜHLMANN fassen diesen Umstand wie folgt zusammen:

„Unter der Einwirkung solcher Institutionen vertraue ich den Mitmenschen, nicht weil ich sie aus persönlichen Erfahrungen kenne, sondern weil ich über Kenntnisse des institutionellen Arrangements ihrer Aktionen verfüge. Institutionen mit einer derartig vertrauensgenerierenden Funktion erklären mir Fremde zu Quasi-Bekannten, denen ich vertrauen kann.“ (Freitag/Bühlmann (2005), S. 582)

Dabei müssen politische Institutionen im Wesentlichen drei Eigenschaften besitzen, um die Entwicklung *generalisierten Vertrauens* zu fördern.

Erstens sollen Institutionen Fairness, Gerechtigkeit, Unbestechlichkeit, Unparteilichkeit und Wahrheit als grundlegende Normen des gesellschaftlichen Zusammenlebens etablieren und durchsetzen. Hierzu müssen sich zunächst die Angehörigen der jeweiligen Institutionen öffentlich und nachprüfbar zu diesen Grundsätzen bekennen sowie Verletzungen der genannten Grundsätze von Institutionen wirksam sanktioniert werden können. Verhalten sich die Institutionen im Sinne der aufgestellten Normen, so wächst unter den Mitgliedern der

3. Theorie: Vertrauen

Gesellschaft die Überzeugung, dass zwischenmenschliches Vertrauen nicht missbraucht wird bzw. ein Missbrauch durch entsprechende Maßnahmen der zuständigen Institutionen geregelt wird. (Freitag/Bühlmann (2005), S. 582; Freitag/Bühlmann (2009), S. 1544) Es soll somit gelten:

„Institutions that are seen as incorruptible, nonpartisan, just, and sanctioners of uncooperative behavior have a greater capacity to promote social trust. Increased individual trust in the judicial and law enforcement systems, a stronger rule of law, a more independent judiciary, and lower levels of corruption increase the likelihood that an individual will develop a high level of generalized trust.“ (Freitag/Bühlmann (2009), S. 1544f)

Zweitens sollen Institutionen universalistisch ausgerichtet sein und Ungleichheiten zwischen den Mitgliedern einer Gesellschaft ausgleichen. Gleichberechtigung sowie eine gerechte Verteilung von Verpflichtungen auf alle Gesellschaftsmitglieder sorgt für einen Anstieg des wechselseitigen Vertrauens, indem die Angst vor einer einseitigen Übervorteilung anderer reduziert wird. *Generalisiertes Vertrauen* ergibt sich somit infolge einer institutionell geförderten Verminderung sozialer Ungleichheit. (Freitag/Bühlmann (2005), S. 582f; Freitag/Bühlmann (2009), S. 1545) So wird folgendes angenommen:

„More universally oriented systems are more likely to promote the development of generalized trust. Greater income equality and increased activity by the state in promoting equal opportunities promote the probability that individuals will develop generalized trust.“ (Freitag/Bühlmann (2009), S. 1545)

Drittens sollen Institutionen Mechanismen zur innergesellschaftlichen Machtteilung sowie zur Lösung von Konflikten bereitstellen. Es steht dabei allerdings eher die Verankerung dieser Mechanismen innerhalb der Gesellschaft als Verhaltensdisposition ihrer Mitglieder im Fokus und weniger die Bereitstellung und Ausführung durch die Institutionen. Eine solche Verankerung von Verhaltensdispositionen kann auf der einen Seite durch positive Erfahrungen mit Institutionen und deren Übertragung auf andere Akteure geschehen. Es muss dafür lediglich vermieden werden, dass durch institutionelle Entscheidungen eine immer wiederkehrende einseitige Bevorteilung bestimmter Individuen oder bestimmter Gruppen entsteht. Institutionen dürfen demnach nicht parteiisch agieren und sollten den Ausgleich zwischen Mehrheiten und Minderheiten fördern. Eine zweite Möglichkeit der Verankerung stellt eine explizite Sozialisation der Gesellschaftsmitglieder in den von den Institutionen vermittelten Normen und Werten dar. (Freitag/Bühlmann (2005), S. 583f; Freitag/Bühlmann (2009), S. 1545ff) Für FREITAG/BÜHLMANN ergibt sich daraus folgende Hypothese:

„Institutions with a greater capacity for consensual and power-sharing are more likely to facilitate the development of social trust. The more transparent power-sharing structures are, the better the chances for minority participation through proportional representation, and the better political rights are protected, the more likely it is that individuals will develop generalized trust.“ (Freitag/Bühlmann (2009), S. 1547)

3.2.4 Ethnische Diversität

Ethnische Heterogenität als mögliche Determinante *generalisierten Vertrauens* stellt ein vergleichsweise neues Feld dar. Grundlegend für die mögliche Einflussnahme ethnischer Heterogenität auf *generalisiertes Vertrauen* ist dabei die Annahme, dass alles, was innerhalb einer Gesellschaft soziale Distanz zwischen den Bürgern schafft, sich in der Regel negativ auf

generalisiertes Vertrauen auswirkt. Entsprechend wird vermutet, dass ethnische Heterogenität zwar förderlich für die Entwicklung intraethnischen Vertrauens ist, dies aber auf Kosten von interethnischem bzw. *generalisiertem Vertrauen* geschieht. Grund hierfür können Abgrenzungstendenzen zwischen den verschiedenen ethnischen Gruppen sein, die durch eine einseitige Intensivierung des Vertrauens verstärkt werden. Daneben kann ethnische Heterogenität auch Auswirkungen auf den Radius des Vertrauens haben und diesen einschränken. So werden weniger Menschen als Teil der eigenen Wertegemeinschaft angesehen, was die Bereitschaft zur Ausdehnung des Vertrauens auf Fremde verringert. Gleichzeitig kann sich auch der Gegenstandsbereich hinsichtlich dessen, dem bisher Vertrauen gewährt wurde, verkleinern. (Nannestad (2008), S. 426ff)

Im Wesentlichen existieren drei widerstreitende theoretische Ansätze, die die Beziehung zwischen ethnischer Heterogenität und *generalisiertem Vertrauen* zu erklären versuchen. Die *Kontakttheorie* geht davon aus, dass ethnische Vielfalt dazu beiträgt Vorurteile abzubauen und wechselseitige Toleranz und Solidarität fördert. In der Folge lösen sich ethnozentristische Einstellungen auf und *ingroup-outgroup*-Unterschiede werden abgeschwächt, sodass sich *outgroup*-Vertrauen entwickeln kann. Entsprechend sollte ethnische Diversität einen positiven Effekt auf *generalisiertes Vertrauen* haben. Die *Konflikttheorie* geht indes davon aus, dass ethnische Diversität Auseinandersetzungen um knappe sozioökonomische Ressourcen zur Folge hat und so die Differenzierung zwischen *ingroup* und *outgroup* befeuert. Je größer die Nähe zu Angehörigen anderer ethnischer Gruppen, desto mehr zieht man sich auf die eigene Gruppe zurück und desto größer wird die Feindschaft gegenüber der *outgroup*. Ethnische Diversität hat nach diesem Verständnis somit einen negativen Effekt auf *generalisiertes Vertrauen*. Die *Konstriktttheorie* postuliert dagegen, dass ethnische Heterogenität sowohl auf das *ingroup*- wie auf das *outgroup*-Vertrauen einen negativen Einfluss ausübt. Eine ethnisch-heterogene Umgebung sorgt somit dafür, dass sich die betroffenen Personen aus dem sozialen Leben zurückziehen und ein sogenanntes „*hunkering down*“ einsetzt. (Putnam (2007), S. 141ff)

PUTNAM verwendet zur Prüfung von Kontakt-, Konflikt- und Konstriktttheorie den *Social Capital Community Benchmark Survey* und findet dabei Hinweise darauf, dass sowohl Konflikt- wie auf Konstriktttheorie eine gewisse Gültigkeit aufweisen. So zeigt sich, dass interethnisches Vertrauen in ethnisch homogenen Gegenden stärker ausgeprägt ist, als dies in ethnisch heterogenen Gegenden der Fall ist. Außerdem vertrauten Menschen, die in ethnisch heterogenen Gemeinden lebten, ihren Nachbarn weniger. Beide Resultate entsprechen den Annahmen der Konflikttheorie und widersprechen den Annahmen der Kontakttheorie. Daneben ist allerdings das Vertrauen in die eigene ethnische Gruppe in einer ethnisch vielfältigen Umgebung ebenfalls niedrig, was wiederum der Konflikttheorie widerspricht aber der Konstriktttheorie entspricht. (Putnam (2007), S. 147ff)

3.2.4.1 Der Einfluss statischer und dynamischer Diversität auf Vertrauen (Gundelach)

Allgemein definiert GUNDELACH soziales Vertrauen „as the expectation that others will contribute to the well-being of a person or a group, or at least that they will refrain from harmful actions“ (Gundelach (2014b), S. 128).

GUNDELACH eröffnet ähnlich wie USLANER ein Kontinuum mit *partikularisiertem* und *generalisiertem Vertrauen* als äußere Pole. Dabei beschreibt *partikularisiertes Vertrauen* ein

3. Theorie: Vertrauen

Vertrauen in Personen aus dem sozialen Nahbereich, zu denen man eine persönliche Beziehung hat, wobei irrelevant ist, welcher sozialen Gruppe sie angehören. Demgegenüber steht mit *generalisiertem Vertrauen* eine Form des Vertrauens, die sich weder auf spezifische Situationen noch auf bestimmte Personen beschränkt, sondern vielmehr eine abstrakte Vertrauenseinstellung gegenüber Menschen im Allgemeinen darstellt. (Gundelach (2014a), S. 27ff; Gundelach (2014b), S. 125f)

Dazwischen angeordnet sind das sogenannte identitätsbasierte/kategoriale Vertrauen und das sogenannte kontextsensitive Vertrauen (*community trust*). Ersteres beschreibt dabei ein Vertrauen zu Personen, mit denen man eine gemeinsame soziale Identität teilt. Grundlage dieser Vertrauensform sind geteilte Normen, Werte und Interessen, die eine geteilte Gruppenzugehörigkeit anzeigen und die wechselseitige Antizipation des zukünftigen Verhaltens der Gruppenmitglieder ermöglichen. So wird davon ausgegangen, dass innerhalb der spezifischen sozialen Gruppe verschiedene Handlungssituationen in ähnlicher Art und Weise beurteilt werden. Auf diese Weise wird es möglich, auch solchen Personen Vertrauen entgegenzubringen, die man zwar nicht persönlich kennt, die man aber als Mitglieder der eigenen sozialen Kategorie identifiziert hat. (Gundelach (2014a), S. 29f)

Kontextsensitives Vertrauen wird von GUNDELACH auch als Nachbarschaftsvertrauen bezeichnet und bezieht sich auf Personen in einem speziellen Kontext, wobei die Informationslage hinsichtlich dieser Personen mehr oder weniger stark elaboriert sein kann. Diese Form des Vertrauens betrifft somit sowohl fremde wie auch bekannte Personen. (Gundelach (2014a), S. 30f)

Hinsichtlich der Wirkung ethnischer Heterogenität auf *generalisiertes Vertrauen* ergeben sich zwei sich zueinander gegenteilig verhaltende Erklärungsansätze. So geht die Kontakttheorie davon aus, dass durch wiederholte und kontinuierliche Begegnungen mit Mitgliedern von Fremdgruppen unter verschiedenen situationellen Rahmenbedingungen Lernprozesse hinsichtlich der Fremdgruppe einsetzen und negative Überzeugungen oder Vorurteile abgebaut werden. Die stetige Konfrontation mit den Einstellungen und Verhaltensweisen der Fremdgruppe erzeugt auf diese Weise Vertrautheit. (Gundelach (2014a), S. 55ff, S. 66; Gundelach (2014b), S. 128; Gundelach/Freitag (2014), S. 1240; Gundelach/Traunmüller (2014), S. 599ff)

Demgegenüber geht die Konflikttheorie davon aus, dass Vertrauen sich vornehmlich in ethnisch homogenen Kontexten entwickelt, innerhalb derer ein Konsens über soziale Normen und angemessenes Verhalten vorherrscht, der wechselseitige Verhaltensantizipationen zulässt und auf diese Weise Vertrautheit erzeugt. In ethnisch heterogenen Gesellschaftskontexten ist dieser *common sense* weniger stark ausgeprägt oder gar nicht vorhanden, was Unsicherheiten im Umgang miteinander und damit auch Misstrauen Vorschub leistet. Moderiert wird dieser Effekt durch eine Eigengruppenfavorisierung und die Bedrohung materieller und ideeller Werte aufgrund gesteigerter Konkurrenz. Die Bevorzugung der eigenen sozialen Gruppe dient dabei im Wesentlichen der Reduktion sozialer Komplexität und ist somit nicht per se negativ zu bewerten. So muss nicht jede Situation des täglichen Lebens im Detail analysiert werden, um erst dann handeln zu können. Der Wettbewerb zwischen ethnischen Gruppen um knappe Ressourcen betrifft dabei vornehmlich ökonomische Ressourcen, politische Macht und die kulturelle Vorherrschaft. (Gundelach (2014a), S. 43ff, S. 64ff; Gundelach (2014b), S. 128f; Gundelach/Freitag (2014), S. 1239f; Gundelach/Traunmüller (2014), S. 601)

Ob *generalisiertes Vertrauen* durch ethnische Heterogenität gefördert wird, hängt nach GUNDELACH im Wesentlichen von deren Ausgestaltung ab. So hat ethnische Heterogenität je

nachdem, ob es sich um eine statische oder dynamische Diversität handelt, einen positiven oder negativen Einfluss auf *generalisiertes Vertrauen*. Statische Diversität beschreibt in diesem Zusammenhang eine ethnische Heterogenität, die ein traditionell stabiles und bereits etabliertes Gesellschaftsmerkmal in der Form darstellt, als dass ethnische Minderheiten bereits lange Zeit Teil der Gemeinschaft sind. Dagegen umreißt dynamische Diversität eine ethnische Heterogenität, die sich aus Migrationsbewegungen in der jüngeren Vergangenheit speist. (Gundelach (2014a), S. 41; Gundelach (2014b), S. 129)

Im Falle von statischer Diversität ist mit einem positiven Effekt auf *generalisiertes Vertrauen* zu rechnen, der mit Hilfe der Kontakttheorie erklärt werden kann. Durch langfristigen und kontinuierlichen Kontakt mit ethnischen Minderheiten konnten sich bereits wechselseitig Automatismen hinsichtlich der Verhaltensantizipation entwickeln und Unsicherheiten und Ängste im Miteinander abgebaut werden. Dynamische Diversität hat dagegen einen negativen Effekt auf *generalisiertes Vertrauen*, was anhand der Konflikttheorie erklärt wird. So erzeugen kürzliche und schnell verlaufende Migrationsbewegungen mehr Ängste und haben ein größeres wahrgenommenes Bedrohungspotenzial. (Gundelach (2014a), S. 64ff; S. 129)

3.2.4.2 Similarität und Homogenität als Verstärker des Vertrauens (Öberg/Oskarsson/Svensson)

ÖBERG/OSKARSSON/SVENSSON unterscheiden zunächst zwei Vertrauenskonzeptionen. So kann Vertrauen in Form einer Disposition auftreten (*disposition to trust*), wobei die Erwartung eines grundlegenden guten Willens seitens anderer Individuen handlungsleitend für die Vertrauensvergabe ist. Menschen werden somit grundsätzlich positive Intentionen unterstellt. Solches Vertrauen ist unabhängig von der Entwicklung persönlicher Beziehungen zwischen den Vertrauensparteien und weist deshalb eine besonders hohe Stabilität auf. Gleichzeitig ist es eng an den Charakter der jeweiligen Person gebunden und lässt die individuelle Vertrauenswürdigkeit anderer Personen unbeachtet. (Öberg/Oskarsson/Svensson (2011), S. 348; Oskarsson/Svensson/Öberg (2007), S. 3; Oskarsson/Svensson/Öberg (2009), S. 172f)

Demgegenüber steht ein relationales Verständnis von Vertrauen (*relational view of trust*). Vertrauen resultiert in dieser Lesart aus Information über die Vertrauenspartner und vergangenen Erfahrungen mit ähnlichen Situationen, die eine Abschätzung des zukünftigen Verhaltens der Vertrauenspartner ermöglichen. Vertrauen wird somit als Bestandteil und Produkt sozialer Beziehungen zwischen zwei oder mehr Individuen angesehen. Dabei existieren zwei Ansätze, den Mechanismus relationalen Vertrauens zu erklären. Auf der einen Seite kann relationales Vertrauen auf der Annahme eines eingebetteten Interesses basieren. Hierbei wird davon ausgegangen, dass der Vertrauenspartner ein Interesse an einem selbst und dem Fortbestand der sozialen Beziehung hat und sich deshalb vertrauenswürdig verhält. Auf der anderen Seite kann sich relationales Vertrauen auch aus der Annahme einer moralischen Verpflichtung des Vertrauenspartners, das in ihn gesetzte Vertrauen nicht zu enttäuschen, entstehen. ÖBERG/OSKARSSON/SVENSSON gehen indes davon aus, dass keine der beiden Ansätze vorzuziehen ist, sondern dass sich relationales Vertrauen vielmehr aus einer Kombination ergibt bzw. beide gleichwertig sind. (Öberg/Oskarsson/Svensson (2011), S. 348f; Oskarsson/Svensson/Öberg (2007), S. 3f; Oskarsson/Svensson/Öberg (2009), S. 173, S. 302)

ÖBERG/OSKARSSON/SVENSSON gehen bei ihrer Beschreibung der Verbindung von ethnischer Diversität bzw. sozialer Heterogenität und Vertrauen von einem relationalen Vertrauenskonzept

3. Theorie: Vertrauen

aus. Strenggenommen handelt es sich bei dem nachfolgend verwendeten Vertrauensbegriff somit nicht um *generalisiertes Vertrauen*, sondern um eine Form *partikularen Vertrauens*. Allerdings ergibt sich aus dem weiter oben beschriebenen Aspekt der moralischen Verpflichtung eine Verbindung zum Konzept des *generalisierten Vertrauens*. Das Verständnis von Vertrauen als Disposition kommt allerdings dem Ansatz *generalisierten Vertrauens* näher.

Nach ÖBERG/OSKARSSON/SVENSSON wird die Entstehung von Vertrauen im Wesentlichen durch individuelle Charakteristika, den institutionellen Kontext sowie ethnische Diversität und soziale Heterogenität beeinflusst. Hinsichtlich individueller Charakteristika spielen insbesondere Machtressourcen, Bildung, Einkommen und individuelle Erfahrungen mit Sicherheit und Unglück eine große Rolle. Der institutionelle Kontext wirkt in Form von unparteiischen, unbestechlichen und gleichheitsfördernden Institutionen positiv auf Vertrauen. (Öberg/Oskarsson/Svensson (2011), S. 347f)

Ethnischer Diversität und sozialer Heterogenität wirken vornehmlich entlang zweier voneinander verschiedenen kontextsensitiven Mechanismen auf Vertrauen, die ÖBERG/OSKARSSON/SVENSSON als *similarity* bzw. *homogeneity hypothesis* bezeichnen. Der Similaritätshypothese zur Folge basiert die Vertrauensentscheidung auf der Abschätzung der eigenen Position in Relation zu anderen Personen innerhalb einer Gemeinschaft. Die eigene Stellung beeinflusst dabei das individuelle Vertrauen in der Form, dass vornehmlich solchen Personen Vertrauen entgegengebracht wird, die einem bekannt sind oder die einem ähnlich sind. Vertrauen ergibt sich dann aus unseren individuellen Einstellungen gegenüber anderen, sodass eher solchen Personen und Fremden vertraut wird, die uns ähnlich sind, als solchen, die keine Similaritäten mit uns aufweisen. Es ergibt sich somit eine negative Korrelation zwischen ethnischer Diversität und Vertrauen. (Öberg/Oskarsson/Svensson (2011), S. 349f)

Kritik an der Erklärungskraft der Similaritätshypothese ergibt sich daraus, dass nun jegliche Art von Ähnlichkeit zwischen Personen und somit auch negativ konnotierte Similarität sich positiv auf Vertrauen auswirken müsste, was aber in der Regel nicht der Fall ist. ÖBERG/OSKARSSON/SVENSSON entwickeln deshalb mit der Homogenitätshypothese eine weitere Perspektive auf den Zusammenhang von ethnischer Diversität bzw. sozialer Heterogenität und Vertrauen, die die Zusammensetzung des jeweiligen Kontextes stärker in den Fokus rückt und weniger die Position des einzelnen Individuums innerhalb des Kontextes betrachtet. Vertrauen hängt hierbei davon ab, wie die Relationen der Akteure innerhalb einer Gemeinschaft wahrgenommen werden. Vertrauen wird durch das kontextuelle Setting beeinflusst und ist dabei unabhängig von der individuellen Position innerhalb dieses Settings. Als Moderator fungiert dabei der Homogenitätsgrad des jeweiligen Kontextes. Unter einem homogenen Kontext wird dabei eine Umgebung verstanden, die ein niedriges Level an sozialen Konflikten aufweist und in der enge soziale Netzwerke vorherrschen. Beides wirkt sich positiv auf Vertrauen aus, auch wenn man selbst nur wenig mit den Individuen innerhalb dieses Kontextes gemeinsam hat. In homogenen Kontexten kann eher davonausgegangen werden, dass Verhaltens- und Gemeinschaftsnormen eingehalten werden bzw. eine Logik der Angemessenheit hinsichtlich sozialer Interaktion eingehalten wird. Auf dieser Basis wird angenommen, dass andere ihr Wort halten werden und sich vertrauenswürdig verhalten werden unabhängig davon, ob sie Ähnlichkeiten hinsichtlich Ethnizität oder sozioökonomischer Faktoren mit einem selbst aufweisen. In einer Umgebung ohne solche Gemeinschaftsnormen oder eine Anleitung für angemessenes Handeln ist Vertrauen dagegen risikoreicher, kann man doch nicht von einer Befolgung allgemeingültiger Regel des Zusammenlebens ausgehen. (Öberg/Oskarsson/Svensson (2011), S. 346, S. 350ff)

3. Theorie: Vertrauen

Zusammenfassend gehen ÖBERG/OSKARSSON/SVENSSON nun von folgenden Verbindungen zwischen Vertrauen und sozialer bzw. ethnischer Heterogenität aus:

„H1: *The similarity hypothesis: Similarity between the individual and the context strengthens trust (and dissimilarity drives down trust).*“ (Öberg/Oskarsson/Svensson (2011), S. 350)

„H2: *The homogeneity hypothesis: Homogeneity fosters trust. This could also be expressed the other way around: Diversity drives trust down.*“ (Öberg/Oskarsson/Svensson (2011), S. 352)

„H3: *The conditional hypothesis: The effect on trust of homogeneity is positive among individuals who are similar to the other persons comprising the context but negative among individuals who are dissimilar to others within their contexts.*“ (Öberg/Oskarsson/Svensson (2011), S. 352)

Wie die letzte der drei Hypothesen zeigt, ergeben sich zwischen Similarität und Homogenität mitunter auch Verstärkereffekte. So führt eine Erhöhung der Homogenität und der Bedingung von Similarität zu einer Stärkung des Vertrauens.

3.2.5 Generalisiertes Vertrauen vs. Partikulares Vertrauen

An dieser Stelle sei zunächst zurück ins Gedächtnis gerufen, dass die Funktion sozialen Vertrauens in erster Linie darin besteht, die mit Interaktionen von Individuen verbundene soziale Komplexität zu mindern, indem vorherrschende Informationsdefizite durch die Annahme ersetzt werden, dass der jeweilige Vertraute ausschließlich wohlwollende Handlungsabsichten hat. Ohne wechselseitiges oder zumindest einseitiges Vertrauen, bei dem einer der Vertrauenspartner in Vorleistung tritt, ergeben sich im Normalfall keine Wechselbeziehungen zwischen Individuen. Erst Vertrauen macht Kooperation möglich. (Preisendörfer (1995), S. 266ff)

Im vorangegangenen Abschnitt wurden im Wesentlichen zwei Vertrauensformen beschrieben, sogenanntes *partikulares* und *generalisiertes Vertrauen*, wobei der Fokus auf der Genese und den Bedingungsfaktoren bzw. Determinanten *generalisierten Vertrauens* lag.

Partikulares Vertrauen beschreibt ein Vertrauen zwischen Personen, die einander kennen, und wird aufgrund dieser Eigenschaften ausschließlich auf der Mikroebene wirksam. Solches Vertrauen ergibt sich im Wesentlichen aus Informationen über die Vertrauten und vergangenen Erfahrungen mit den Vertrauten. Es wird dabei auch als *thick, personal* oder *specific trust* bezeichnet und entspricht somit im Grunde dem, was landläufig unter Vertrauen verstanden wird. Beispiele für ein solches Vertrauen sind das Verleihen eines wertvollen Gegenstandes an einen Verwandten oder Freund, die Übertragung einer wichtigen Aufgabe an einen Kollegen oder Untergebenen oder die Bitte einer Person aus dem sozialen Nahbereich um einen Gefallen. In all diesen Fällen wird auf Basis von erfahrungsgeneriertem Wissen darauf vertraut, dass der oder die Vertraute(n) sowohl die Kompetenz als auch die Intention besitzt, im Interesse des Vertrauenden zu handeln. Entsprechend folgt aus der vorangegangenen Auflistung und den Erörterungen der verschiedenen Vertrauensansätze und ihrer konzeptionellen Theoriefamilien folgende Definition für Partikulares Vertrauen¹⁹:

¹⁹ Wird im weiteren Verlauf dieser Studie von „Partikularem Vertrauen“ gesprochen, so beziehen sich die Ausführungen auf die an dieser Stelle formulierte Definition. Andere Schreibweisen (z.B. partikulares Vertrauen, particularized trust etc.) sind anderen Vertrauenskonzeptionen zuzuschreiben.

Partikulares Vertrauen stellt eine dreiwertige Relation dar, die stets auf spezifische Vertrauenssubjekte und spezifische Vertrauensobjekte beschränkt ist. Das heißt: Der Vertrauende vertraut dem Vertrauten bezüglich eines bestimmten Vertrauensgegenstandes bzw. der Vertrauende vertraut auf ein bestimmtes Verhalten des Vertrauten.

Partikulares Vertrauen ergibt sich auf der Basis eines unvollständigen Wissens um die individuelle Vertrauenswürdigkeit und Zuverlässigkeit des Vertrauten und ist auf bekannte Individuen gerichtet. Es nimmt an, dass der Vertraute die Fähigkeit besitzt, das von ihm erwartete Handeln umzusetzen, und motiviert ist, im Interesse des Vertrauenden zu handeln. Der Vertrauende akzeptiert dabei sowohl die Abhängigkeit vom Verhalten des Vertrauten beim Erreichen eines speziellen Nutzens, der ohne die Unterstützung des Vertrauten nicht realisiert werden könnte, als auch das hieraus resultierende Risiko der Verletzbarkeit seiner Person. Der Vertrauende hat die Erwartung, dass der Vertraute ein spezielles für ihn vorteilhaftes Verhalten zeigen wird, ohne dass er den Vertrauten dahingehend überwachen müsste.

Partikulares Vertrauen entwickelt sich in der Regel innerhalb andauernder Beziehungen, die für alle beteiligten Parteien gewinnbringend sind. Je länger eine solche Beziehung andauert, umso gewinnbringender und nützlicher wird sie und umso vertrauensvoller und vertrauenswürdiger werden die beteiligten Personen. Die Annahme eines wohlwollenden Verhaltens seitens des Vertrauten und einer Vermeidung eines Vertrauensmissbrauchs seinerseits, ist deshalb nicht zuletzt durch ein Interesse an einer Fortsetzung der Beziehung begründet.

Partikulares Vertrauen kann einseitiger oder wechselseitiger Natur sein. Unter der Bedingung der Einseitigkeit bringt der Vertrauende dem Vertrauten Vertrauen entgegen, während unter der Bedingung der Wechselseitigkeit beide an der Vertrauensbeziehung beteiligten Parteien sowohl die Rolle des Vertrauenden als auch die Rolle des Vertrauten innehaben.

Während Partikulares Vertrauen auf bekannte Personen aus dem nahen sozialen Umfeld bezogen ist, d.h. auf Personen, über die wir Informationen hinsichtlich ihrer Vertrauenswürdigkeit besitzen oder aus ihren Handlungen ableiten können, sind wir im Alltag in einer Vielzahl von Situationen mit Personen konfrontiert bzw. gezwungen zu interagieren, über die wir keine expliziten Informationen hinsichtlich ihrer Vertrauenswürdigkeit besitzen. Aufgrund der Flüchtigkeit des Kontaktes mit den meisten dieser Menschen lässt sich allerdings keine Vertrauensbeziehung im Sinne eines Partikularen Vertrauens aufbauen. Zudem würde sich Vertrauen dieser Art in solchen Alltagssituationen keinesfalls als funktional erweisen, da es eher eine Eskalation, denn eine Reduktion sozialer Komplexität nach sich ziehen würde. So wäre ein ständiges Abwägen, der mit der Vertrauensvergabe verbundenen Risiken vor dem Hintergrund der individuellen Spezifika des Vertrauten und des jeweiligen Vertrauensobjektes notwendig. Zusätzlich trägt die fortschreitende Ausdifferenzierung der Gesellschaft zur Entwicklung von Expertensystemen bei, in deren Folge eine Professionalisierung von Wissensbeständen stattfindet. Daraus ergibt sich, dass die Mitglieder der Gesellschaft auf einer Vielzahl relevanter Sachgebiete nur über Laienwissen oder kein adäquates Wissen verfügen, sondern bei der Bewertung von Sachverhalten auf entsprechendes Expertenwissen angewiesen sind. Charakteristisch für solches Expertenwissen ist allerdings, dass es oft nicht ohne Weiteres nachprüfbar ist bzw. eine Überprüfung spezielle Kenntnisse oder Ressourcen erfordert. Außerdem erweist sich eine Kontrolle meist als wenig ökonomisch, weshalb man notwendigerweise darauf vertrauen muss, dass die Einschätzungen und Empfehlungen der

3. Theorie: Vertrauen

Experten korrekt sind. Mit anderen Worten: Das Konzept Partikularen Vertrauens kommt an dieser Stelle an seine Grenzen.

Das Problem des Informationsdefizites gegenüber fremden Personen lässt sich indes nur durch eine generalisierte Form des Vertrauens lösen, bei der Vertrauensentscheidungen einer Heuristik folgen. Vertrauen dieser Art wird auch *thin*, *impersonal* oder *diffuse trust* genannt und wird insbesondere dann wirksam, wenn es um einmalige oder flüchtige Kontakte geht. Beispiele hierfür sind das Vertrauen in die Fähigkeiten und die Kompetenz von Piloten, Busfahrern oder Taxifahrern. Bei der Nutzung der entsprechenden Verkehrsmittel muss jedes Mal darauf vertraut werden, dass der jeweilige Lenker in der Lage ist das Flug- bzw. Fahrzeug zu steuern, nicht unter Drogen steht, suizidale Tendenzen aufweist etc.. Eine umfassende Bestimmung der Charakterdispositionen dieser Personen ist in der Regel aufgrund der zeitlichen Limitation der Interaktion und den Einschränkungen hinsichtlich einer Kontaktaufnahme nicht möglich, sodass kein kalkuliertes Vertrauen im Sinne eines Partikularen Vertrauens etabliert werden kann. Eine Überbrückung der aus diesen Beziehungskonstellationen resultierenden sozialen Komplexität lässt sich nur durch ein Generalisiertes Vertrauen bewerkstelligen.

Die Notwendigkeit eines Generalisierten Vertrauens infolge des Kontrastes von Laien- und Expertenwissen lässt sich beispielweise anhand der Arbeit von Journalisten, Wissenschaftlern und Ärzten veranschaulichen. Die Aufgabe von Journalisten besteht im Normalfall in der Verdichtung und Aufbereitung von Informationen, da es für den Normalbürger kaum praktikabel ist, selbstständig sämtliche relevanten Daten insbesondere zu komplexen Thematiken zu erfassen. Die Adressaten von Journalisten vertrauen dabei darauf, dass die Dokumentationen von Journalisten nach Möglichkeit Resultat einer gewissenhaften Recherche sind und keine Falschinformationen bzw. Fehlinterpretationen, sondern ausschließlich Fakten enthalten. Ähnlich verhält es sich mit den Forschungsergebnissen von Wissenschaftlern. Der Laie muss hier darauf vertrauen, dass diese korrekt sind und durch den Prozess des Peer-Review validiert werden. Er selbst ist mangels Fachwissen nicht dazu in der Lage, sondern muss sich auf entsprechende Kontrollinstanzen und die Aufrichtigkeit bzw. Tugendhaftigkeit des Wissenschaftlers verlassen. Hinsichtlich eines Arztes kommt Generalisiertes Vertrauen dann zum Tragen, wenn es um die Verlässlichkeit von Diagnosen geht. Der Laie muss darauf vertrauen, dass der Arzt kompetent ist eine Krankheit zu erkennen und die richtigen Maßnahmen zu deren Behandlung zu ergreifen. Dem Laien fehlt hier im Allgemeinen das notwendige Expertenwissen, um eine eigenständige Überprüfung zu Wege zu bringen.

Daneben spielt Generalisiertes Vertrauen auch eine zentrale Rolle bei der Anbahnung neuer Beziehungen bzw. für das Kennenlernen fremder Personen. So verfügen wir hinsichtlich uns unbekannter Personen in der Regel über keinerlei Informationen. Informationen über unser Gegenüber stellen allerdings eine der Grundvoraussetzungen dafür dar, einzuschätzen, ob wir beispielsweise jemanden sympathisch finden oder jemand eine Bedrohung für uns darstellt. Es ist uns unmöglich, das zukünftige Verhalten des Anderen zu antizipieren. Vor diesem Hintergrund stellt sich dann allerdings die Frage, warum wir trotzdem dazu in der Lage sind, neue Beziehungen mit uns bisher vollkommen unbekannten Personen einzugehen, wenn dafür doch ein Mindestmaß an Vertrauen notwendig ist. Da Partikulares Vertrauen auf Kenntnissen hinsichtlich des Vertrauten basiert, kann dieses nicht dieses Minimum an Vertrauen liefern. Generalisiertes Vertrauen fungiert an dieser Stelle als sogenannter *leap of faith*, der dieses Minimalvertrauen bereitstellt. Man tritt somit der fremden Person gegenüber in gewisser Weise in Vorleistung, indem man ihr wohlwollend entgegentritt, bis man mehr Informationen gesammelt hat, um zu einem abschließenden Urteil zu kommen. Ein Beispiel für eine solche

3. Theorie: Vertrauen

Situation ergibt sich, wenn Eltern ihr Kind in die Obhut eines neuen Babysitters geben. Sie wissen noch nicht, ob der Babysitter das eigene Kind gut behandeln wird, vertrauen aber darauf.

Die nachfolgende Definition Generalisierten Vertrauens ergibt sich dabei als Kondensat der vorausgegangenen theoretischen Erörterungen²⁰:

Generalisiertes Vertrauen beschreibt eine prinzipiell stabile, in ihrem Ausmaß allerdings variable, individuelle Charakterdisposition, die eine abstrakte Vertrauenseinstellung gegenüber Menschen im Allgemeinen bzw. generalisierten Anderen ohne Beschränkung auf spezifische Situationen oder Personen beinhaltet. Als solche ist sie unabhängig von der Entwicklung einer persönlichen Beziehung zum Vertrauenspartner und lässt die individuelle Vertrauenswürdigkeit anderer Personen unberücksichtigt.

Generalisiertes Vertrauen geht somit davon aus, dass den meisten Menschen vertraut werden kann, ohne dass spezifische Informationen über diese verfügbar sein müssen. Es wird auf diese Weise auch jenseits von face-to-face-Beziehungen, wie Verwandtschaft, Freundschaft oder Bekanntschaft wirksam und schließt Fremde bzw. Menschen, die einem persönlich nicht bekannt sind, sowie unbekannte Gruppen mit ein.

Ein hohes Maß an Generalisiertem Vertrauen ist vornehmlich durch eine optimistische Weltanschauung, die Wahrnehmung, Herr über das eigene Schicksal zu sein, und die Überzeugung, dass alle Menschen Teil einer gemeinsamen Wertegemeinschaft sind, innerhalb derer fundamentale moralische Werte geteilt werden, gekennzeichnet. Gleichzeitig geht Generalisiertes Vertrauen von einem Prinzip generalisierter Reziprozität aus, wonach für einen erwiesenen Gefallen keine unmittelbare Gegenleistung erwartet wird, sondern darauf vertraut wird, dass in der Zukunft der Gefallen, evtl. auch durch dritte, an der Vertrauensbeziehung gänzlich unbeteiligte Personen, vergolten wird. Determinanten und Bezugsrahmen Generalisierten Vertrauens sind dabei gleichermaßen die soziale Ordnung und politische Verfasstheit einer Gesellschaft, die den Spielraum für angemessenes Handeln innerhalb der sozialen Gemeinschaft definieren.

Generalisiertes Vertrauen dient als sozialer Schmierstoff, der soziale Interaktion und Kooperation in größeren Gemeinschaften und zwischen einander fremden Menschen ermöglicht, indem es die Opportunitätskosten verringert. Hierbei steht insbesondere das einmalige Zusammentreffen von Personen ohne wiederholten Kontakt, die Etablierung neuer Beziehungen mit bisher unbekannten Personen und die Interaktion mit einer großen Anzahl von Personen im Fokus. Es impliziert eine grundsätzliche potentielle Bereitschaft zu wechselseitiger Kooperation und zu Engagement in gemeinsamen Unternehmungen.

²⁰ Wird im weiteren Verlauf dieser Studie von „Generalisiertem Vertrauen“ gesprochen, so beziehen sich die Ausführungen auf die an dieser Stelle formulierte Definition. Andere Schreibweisen (z.B. generalisiertes Vertrauen, generalized trust etc.) sind anderen Vertrauenskonzeptionen zuzuschreiben.

4. Messung generalisierten Vertrauens

Zur Erfassung bzw. Abbildung *generalisierten Vertrauens* liegen eine Vielzahl von Skalen vor. Nachfolgend sollen diese vorgestellt und diskutiert werden. Begonnen wird dabei mit dem Standardinstrument (*Faith in People-Scale* bzw. *Trust in People-Scale*), welches weltweit in großen survey-gestützten Befragungen zum Einsatz kommt. Daran anschließend werden die gängigen Alternativskalen vorgestellt.

Die Messung *generalisierten Vertrauens* mit Hilfe des Standardinstrumentes sieht sich außerdem mit einer Reihe von methodischen und inhaltlichen Problemen (Existenz *generalisierten Vertrauens*, Vertrauensradius, Vertrauenslevel, reaktive Effekte Multidimensionalität der Standarditems, Differenziertheit der Vertrauensmessung) konfrontiert, die die Güte der Messungen negativ beeinflussen können.

4.1 Standarditems zur Messung *generalisierten Vertrauens*

Für die Messung von *generalisiertem Vertrauen* innerhalb von Fragebogen wird in der Regel eine Reihe von Standarditems verwendet. Das ursprüngliche Standarditem zur Vertrauensmessung wurde 1948 von NOELLE-NEUMANN entwickelt und lautete (Delhey/Newton/Welzel (2011), S. 787; Uslaner (2012b), S. 72; Zmerli/Newton (2008), S. 709; Zmerli/Newton/Montero (2007), S. 39):

„Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?“

1956 erweiterte ROSENBERG das Standarditem um zusätzliche Fragen zur *Faith in People-Scale*. Die verwendeten Items folgen dabei der Logik einer Guttman-Skala und lauten wie folgt (Robinson/Shaver/Wrightsmann (1991), S. 404ff; Rosenberg (1956), S. 690; Rosenberg (1957a), S. 26; Rosenberg (1957b), S. 341; Rosenberg (1965), S. 180f):

4. Messung generalisierten Vertrauens

Abb. 4.1: Faith in People-Scale

1. *Some people say that most can be trusted. Others say you can't be too careful in your dealings with people. How do you feel about it?*

() *Most people can be trusted.*

() *You can't be too careful.*

2. *Would you say that most people are more inclined to help others or more inclined to look out for themselves?*

() *To help others.*

() *To look out for themselves.*

3. *If you don't watch yourself, people will take advantage of you.*

() *Agree*

() *Disagree*

4. *No one is going to care much what happens to you, when you get right down to it.*

() *Agree*

() *Disagree*

5. *Human nature is fundamentally cooperative.*

() *Agree*

() *Disagree*

- Eichstichprobe: N=4585 college students (nationwide sample)
- Interne Konsistenz: $\alpha = .92$
- Konvergente Validität: Wird aus dem Zusammenhang von Berufswahl und Skalenscore der Probanden abgeleitet. Probanden, die soziale Tätigkeiten ausübten, erreichten höhere Werte als Probanden, die eine absatzorientierte Tätigkeit ausübten.

(Quelle: Robinson/Shaver/Wrightsmann (1991), S. 404ff)

Positive Antworten zeigen hierbei die Abwesenheit von Vertrauen an, während negative Antworten die Anwesenheit von Vertrauen anzeigen.

In großen nationalen und internationalen Bevölkerungsumfragen, wie dem WVS (World Value Survey), dem GSS (General Social Survey), dem ESS (European Social Survey), der ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) oder dem SOEP (Sozio-ökonomisches Panel), kommt heutzutage eine verkürzte Version der von ROSENBERG entwickelten Skala zum Einsatz, die sogenannte *Trust in People-Scale*²¹ (Robinson/Shaver/Wrightsmann (1991), S. 406ff):

Abb. 4.2: Trust in People-Scale

1. *Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?*

() *Most people can be trusted.*

() *Can't be too careful*

2. *Would you say that most of the time, people try to be helpful, or that they are mostly just looking out for themselves?*

() *Try to be helpful*

() *Look out for themselves*

3. *Do you think that most people would try to take advantage of you if they got the chance or would they try to be fair?*

() *Take advantage*

() *Try to be fair*

- Eichstichprobe: N=1450 (postelection study 1964); N=1330 (postelection study 1968)
- Inter-Item Korrelationen als Substitut für innere Konsistenz: Die einzelnen Skalenskomponenten interkorrelieren mit $r = .48-.54$ (sowohl in der Studie von 1964 als auch in der Studie von 1968)
- Keine explizite Überprüfung der Validität.

(Quelle: Robinson/Shaver/Wrightsmann (1991), S. 406ff)

²¹ Vgl. deutsche Übersetzung und Adaption der Standarditems (Kunz (2004), S. 204f):

1. „Ganz allgemein gesprochen: Glauben Sie, dass man den meisten Menschen vertrauen kann, oder dass man im Umgang mit anderen Menschen nicht vorsichtig genug sein kann?“
2. „Glauben Sie, dass die meisten Menschen versuchen Sie auszunutzen, wenn Sie die Gelegenheit dazu haben, oder versuchen die meisten Menschen, sich fair zu verhalten?“
3. „Und glauben Sie, dass die Menschen versuchen hilfsbereit zu sein, oder dass die Menschen meistens auf den eigenen Vorteil bedacht sind?“

4. Messung generalisierten Vertrauens

Streng genommen enthält die Skala dabei neben einer Vertrauenskomponente auch eine Hilfsbereitschaftskomponente und eine Fairnesskomponente, sodass sie insgesamt drei verschiedene, allerdings miteinander korrespondierende, Konstrukte erfasst.

Innerhalb der genannten Surveys werden die Vertrauensitems auf teilweise höchst unterschiedliche Art und Weise angewendet. So unterscheiden sich die internationalen Bevölkerungsumfragen WVS (World Values Survey), GSS (General Social Survey) und ESS (European Social Survey) in der Anzahl der verwendeten Vertrauensitems sowie der jeweils angelegten Antwortskala. Im WVS findet lediglich das bereits von NOELLE-NEUMANN entwickelte Standarditem Anwendung, welches mit den oben dargestellten dichotomen Antwortmöglichkeiten bewertet werden kann. Dagegen sind im GSS bereits alle Items der *Trust in People-Scale* im Einsatz. Die jeweils dichotome Antwortskala wird dort aber um die Mittelkategorie „*Depends*“ erweitert. Im ESS werden ebenfalls alle drei Standarditems verwendet. Allerdings erfolgt die Bewertung der Statements hier auf einer 11stufigen Ratingskala. (Reeskens/Hooghe (2008), S. 518ff)

Der ALLBUS (Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften) nutzt die deutsche Übersetzung des ersten Items der *Faith in People-Scale*, wobei die Antwortskala derjenigen aus dem GSS entspricht:

Abb. 4.3: ALLBUS

„Manche Leute sagen, dass man den meisten Menschen trauen kann. Andere meinen, dass man nicht vorsichtig genug sein kann im Umgang mit anderen Menschen. Was ist ihre Meinung dazu?“

- () Den meisten Menschen kann man trauen.
- () Man kann nicht vorsichtig genug sein.
- () Das kommt darauf an.
- () Sonstiges, und zwar...
- () Weiß nicht
- () KA

(Quelle: ALLBUS 2016. Fragebogendokumentation., S. 8)

Das SOEP (Sozio-ökonomisches Panel) verwendet insgesamt sieben Items zur Erfassung von *generalisiertem Vertrauen*. Es handelt sich hierbei im Wesentlichen um die deutsche Übersetzung der *Trust in People-Scale*, wobei hier der Mehrdimensionalität²² bzw. der sogenannten „*double-barreledness*“ der Standarditems Rechnung getragen wird. Um dem Radiusproblem²³ der Messung *generalisierten Vertrauens* zu begegnen, wurde außerdem ein weiteres Item ergänzt, dass explizit das Vertrauen zu Fremden abfragt. (Naef/Schupp (2009), S. 6f) Es ergibt sich für das SOEP nun die folgende Itemstruktur:

²² Das Problem der Mehrdimensionalität der Standarditems wird im Abschnitt „4.3.5 Multidimensionalität der Standarditems“ diskutiert.

²³ Das Radiusproblem der Vertrauensmessung wird im Abschnitt „4.3.2 Radius *generalisierten Vertrauens*“ behandelt.

4. Messung generalisierten Vertrauens

Abb. 4.4: SOEPtrust

Was ist Ihre Meinung zu den folgenden drei Aussagen?

„Im Allgemeinen kann man den Menschen vertrauen.“

„Heutzutage kann man sich auf niemanden mehr verlassen.“

„Wenn man mit Fremden zu tun hat, ist es besser, vorsichtig zu sein, bevor man ihnen vertraut.“

(Die jeweilige Antwortskala lautet: „Stimme voll zu“, „Stimme eher zu“, „Lehne eher ab“, „Lehne voll ab“)

Glauben Sie, dass die meisten Leute...

() Sie ausnützen würden, falls Sie die Möglichkeit dazu hätten...

() oder versuchen würden Ihnen gegenüber fair zu sein?“

Glauben Sie, dass die meisten Leute...

() Sie ausnützen würden, falls Sie die Möglichkeit dazu hätten...

() oder versuchen würden Ihnen gegenüber fair zu sein?“

Würden Sie sagen, dass die Leute die meiste Zeit...

() versuchen, hilfsbereit zu sein...

() oder nur ihre eigenen Interessen verfolgen?“

(Quelle: SOEP 2013 – Erhebungsinstrumente 2013 (Welle 30) des Sozio-oekonomischen Panels: Personenfragebogen, Altstichproben, SOEP Survey Papers, No. 180., S. 4)

Neben den hier vorgestellten Standarditems existiert noch eine Vielzahl weiterer Skalen, die zur Erfassung *generalisierten Vertrauens* verwendet werden. Nachfolgend sollen nun zunächst diese Skalen diskutiert werden, bevor die Probleme der Vertrauensmessung mit den Standarditems thematisiert werden.

4.2 Alternative Vertrauensskalen

Neben der *Faith in People-Scale* bzw. den Standarditems der Vertrauensmessung existiert noch eine Reihe weiterer Skalen, die zur Erfassung *generalisierten Vertrauens* verwendet werden.

Abb. 4.5: Reliabilität und Validität der alternativen Vertrauensskalen

Interpersonal Trust Scale (ITS) (Rotter (1967))

Eichstichprobe: N=4605 introductory psychology students (♀ =299, ♂=248)

Split-Half-Reliabilität: $\rho_{tt}=.76$ bzw. $\rho_{tt}=.75$ (♀) u. $\rho_{tt}=.77$ (♂)

Philosophies of Human Nature (PHN) (Wrightsman (1974))

Eichstichprobe: N=530 Undergraduates (♀ =277, ♂=253)

Split-Half-Reliabilität: $\rho_{tt}=.74$

Konvergente Validität: $r_{tc}=.77$ (Rosenberg's faith in people), $r_{tc}=-.68$ (Machiavellianism)

Machiavellianism (Mach IV) (Christie/Geis (1970))

Eichstichprobe: N=1700 College Students

Split-Half-Reliabilität: $\rho_{tt}=.79$

Konvergente Validität: $r_{tc}=-.70$ bis $-.80$ (PHN)

The Trust Inventory (Couch/Adams/Jones (1996))

Eichstichprobe: N=175 (Undergraduate College Students)

Interne Konsistenz: $\alpha=.91$

Konvergente Validität: $r_{tc}=.66$ (ITS), $r_{tc}=.57$ (PHN)

Inventar zur Erfassung des interpersonellen Vertrauens (IIV) (Kassebaum (2004))

Eichstichprobe: N=320 (15-84 Jahre)

Skala Allgemeines Vertrauen:

Interne Konsistenz: $\alpha=.87$; Split-Half-Reliabilität: $\rho_{tt}=.82$ bis $.89$

Konvergente Validität: $r_{tc}=.58$ bis $.59$ (ITS)

KUSIV3 (Beierlein/Kemper/Kovaleva/Rammstedt (2012))

Eichstichprobe: $N_{1.1}=539$ (Stichprobe 1, Welle 1) bzw. $N_{1.2}=338$ (Stichprobe 1, Welle 2) und $N_3=1134$ (Stichprobe 3)

Interne Konsistenz: $\omega=.85$ (N_1 und N_3)

Retest-Reliabilität: $r_{tt}=.57$ ($N_{1.1}$ und $N_{1.2}$)

Validität: Ableitung aus Zusammenhang mit bekannten typischen Korrelaten

(Quelle: Beierlein/Kemper/Kovaleva/Rammstedt (2012), S. 12ff; Kassebaum (2004), S. 116, S. 119, S. 222ff; Couch/Adams/Jones (1996), S. 312, S. 314; Robinson/Shaver/Wrightsman (1991), S. 376ff, S. 385ff, S. 392ff)

Die Übersicht über die Skalenkennwerte zeigt, dass alle der genannten Skalen eine ausreichende Reliabilität aufweisen und dazu in der Lage sind, *generalisiertes Vertrauen* valide abbilden zu können. Nachfolgend sollen nun die inhaltlichen Spezifika der Vertrauensskalen vorgestellt werden.

4.2.1 Interpersonal Trust Scale (ITS)

ROTTER entwickelte die *Interpersonal Trust Scale* (ITS) zur Erfassung genereller Vertrauens- bzw. Misstrauenserwartungen. Im Speziellen sollte festgestellt werden, inwiefern sich eine Person auf die verbalen oder schriftlichen Äußerungen eines anderen Individuums bzw. einer Gruppe verlässt. Die in die Skala einbezogenen Items betreffen dabei das Ausmaß interpersonales Vertrauens in verschiedenen Situationen und gegenüber unterschiedlichen sozialen Objekten. Im Einzelnen handelt es sich dabei um das Vertrauen gegenüber den Eltern, Verkaufspersonal, der Justiz, Menschen im Allgemeinen, Politikern, Handwerkern, Ärzten, Lehrern,

4. Messung generalisierten Vertrauens

Klassenkameraden und Freunden sowie den Medien. Die meisten der 25 Items befassen sich mit der Glaubwürdigkeit der genannten sozialen Objekte, während andere sich auf den Optimismus bezüglich der zukünftigen gesellschaftlichen Entwicklung im Allgemeinen beziehen. Von diesen Items wurden 12 positiv formuliert, sodass Zustimmung Vertrauen anzeigte, und 13 negativ formuliert, sodass Ablehnung Vertrauen anzeigte. Zusätzlich wurden 15 sogenannte Füllitems verwendet, die die Messintention der Skala verschleiern sollten. Jedes der Statements konnte anhand von fünf Antwortkategorien bewertet werden, wobei die jeweiligen Antwortoptionen mit „strongly agree“ (1), „mildly agree“ (2), „agree and disagree equally“ (3), „mildly disagree“ (4) sowie „strongly disagree“ (5) benannt wurden. Die ITS ist als additive Skala konzipiert, sodass die Summe über die einzelnen Itembewertungen das Ausmaß des *generalisierten Vertrauens* der jeweiligen Befragten anzeigt. Die Vertrauenswerte bewegen sich daher zwischen 25 (geringstes Vertrauen) und 125 (höchstes Vertrauen), wobei ein Wert von 75 Indifferenz anzeigt. (Amelang/Gold/Külbel (1984), S. 199f; Robinson/Shaver/Wrightsman (1991), S. 393ff; Rotter (1967), S. 653ff; Rotter (1971), S. 445f; Rotter (1981), S. 23f)

Faktorenanalytische Untersuchungen innerhalb verschiedener Studien zeigen, dass es sich bei dem mit der ITS erfassten generalisierten Vertrauen keinesfalls um ein eindimensionales Konstrukt handelt, sondern *generalisiertes Vertrauen* sich entweder aus unterschiedlichen Aspekten zusammensetzt oder verschiedene Formen des Vertrauens gemessen werden. Im Wesentlichen wird kritisiert, dass diese Multidimensionalität der durch die Vertrauensdefinition ROTTERs postulierten Eindimensionalität widerspricht. (Chun/Campbell (1975), S. 1058f, S. 1066f; Kaplan (1973), S. 13f; Kassebaum (2004), S. 24)

So findet KAPLAN drei Faktoren, die er mit *institutional trust*, *sincerity* und *caution* benennt. *Institutional trust* betrifft dabei Vertrauen gegenüber den einflussreichsten sozialen Agenten innerhalb der Gesellschaft, während sich *sincerity* auf die Wahrnehmung von Aufrichtigkeit bei anderen bezieht. *Caution* beschreibt schließlich die Angst von anderen ausgenutzt zu werden. (Kaplan (1973), S. 13f)

CHUN/CAMBELL untergliedern die ITS in insgesamt vier Dimensionen. *Political cynicism* fokussiert dabei auf die Skepsis gegenüber der Politik und ihrer Vertreter, während *interpersonal exploitation* sich auf den Selbstschutz und die Vorsicht von als ausbeuterisch und egozentrisch wahrgenommenen Anderen bezieht. *Societal hypocrisy* nimmt auf der einen Seite Bezug auf die Scheinheiligkeit innerhalb der Gesellschaft und thematisiert auf der anderen Seite das Unvermögen sozialer Vorbilder allgemeinen Erwartungen zu entsprechen. *Reliable role-performance* greift schließlich ebenfalls die Scheinheiligkeit der Gesellschaft auf und befasst sich mit der mangelhaften Rollenperformance der übrigen Gesellschaftsmitglieder. (Chun/Campbell (1975), S. 1063f)

WRIGHT/TEDESCHI unterscheiden drei Faktoren, *political trust*, *paternal trust* und *trust of strangers*. *Political trust* beinhaltet dabei Items, die sich auf die Glaubwürdigkeit von Politik, Medien und sportlichen Wettkämpfen beziehen. *Paternal trust* beinhaltet Items, die sich auf die Verlässlichkeit wohlwollender Autoritäten, wie Eltern, Experten, Idealisten, Verkäufer oder Umfrageteilnehmer, beziehen. *Trust of strangers* misst schließlich, als wie wahrscheinlich es die Befragten ansehen, dass unbekannte Individuen sie ausnutzen würden. Zusätzlich betrifft diese Dimension die Annahme, dass andere Menschen egoistisch handeln und ihnen deshalb mit Vorsicht begegnet werden muss. Dabei unterscheiden sich die Items, die auf die jeweiligen Faktoren laden neben der inhaltlichen Ausrichtung auch in ihrer Polung. So sind die Items der Dimensionen *political trust* und *trust of strangers* negativ formuliert, während die Items der Dimension *paternal trust* positiv formuliert sind. Darüber hinaus bestehen zwischen Items, die

4. Messung generalisierten Vertrauens

auf *political trust* und *paternal trust* laden sowie zwischen Items, die auf *trust of strangers* und *paternal trust* laden, Überschneidungen. Diese Verbindung von *paternal trust* mit den beiden anderen Dimensionen werten WRIGHT/TEDESCHI als Hinweis darauf, dass *paternal trust* einen früh erlernten und fundamentalen Aspekt des Vertrauens darstellt. Eine Person, die wohlwollend eingestellten Autoritäten misstrauisch gegenübersteht, wird eher als verhaltensgestört angesehen als eine Person, die Fremden oder Politikern nicht vertraut. (Wright/Tedeschi (1975), S. 471f)

KRAPPEN/VIEBIG/WALTER arbeiten mit einer deutschen Übersetzung der ITS und ergänzen diese um weitere fünf Items. Items, die hoch auf mehrere Faktoren luden, wurden eliminiert. Sie extrahieren insgesamt drei Faktoren, die soziales Misstrauen und soziale Angst gegenüber Fremden, Vertrauen in die Zuverlässigkeit, Ehrlichkeit und Glaubwürdigkeit anderer Menschen sowie mediales Misstrauen bzw. Skepsis gegenüber Informationen aus den Medien abbilden. (Krappen/Viebig/Walter (1982), S. 242f, S. 246)

AMELANG/GOLD/KÜLBEL prüfen eine deutsche Übersetzung der ITS und finden vier Faktoren. Im Einzelnen betreffen diese öffentliche Institutionen und deren Transparenz, Mitmenschen und die von diesen ausgehende Bedrohung, Experten und deren Verhalten sowie die Differenz zwischen Verbal- und Realverhalten. (Amelang/Gold/Külbel (1984), S. 204)

PETERMAN schließlich unterteilt die ITS in vier Dimensionen: „Aussagen über die Gesellschaft und Zukunft im Allgemeinen, Aussagen über politische und soziale Institutionen, Fragen zur Glaubwürdigkeit von Medien und Fragen zur Glaubwürdigkeit einer Vielzahl von Gruppen“ (Petermann (2013), S. 24).

Aufgrund der aufgezeigten Multidimensionalität der *Interpersonal Trust Scale* plädieren KAPLAN, CHUN/CAMPBELL und WRIGHT/TEDESCHI darauf, in zukünftigen Studien ein größeres Augenmerk auf diese Subskalen zu legen. So sollte weniger *generalisiertes Vertrauen* als die verschiedenen Komponenten dieses Vertrauens im Fokus stehen. Außerdem könnten die besonders prägnanten Items in Form einer Kurzskala für *generalisiertes Vertrauen* zum Einsatz kommen. (Chun/Campbell (1974), S. 1069; Kaplan (1973), S. 14; Wright/Tedeschi (1975), S. 476)

Schon bei ROTTER zeigte sich, dass es sich bei *generalisiertem Vertrauen* um ein Konstrukt handelt, welches durch sozial erwünschtes Antwortverhalten verzerrt sein kann. So bearbeiteten die Probanden im Rahmen der Skalenentwicklung neben den Vertrauensitems auch die *Marlowe-Crowne-Social-Desirability-Scale*, wobei sich signifikante Korrelationen zwischen den beiden Skalen ergaben. *Generalisiertes Vertrauen* kann somit als ein durch soziale Erwünschtheit beeinflusstes Konstrukt betrachtet werden. Dieses Ergebnis konnte von nachfolgenden Studien, in denen die ITS verwendet wurde, repliziert werden. So finden WRIGHT/TEDESCHI, dass insbesondere Items, die dem Faktor *paternal trust* zugerechnet wurden, mit der *Marlowe-Crowne-Scale* korrelierten. CHUN/CAMPBELL kritisieren zwar die Ergebnisse von ROTTER hinsichtlich des Einflusses sozial erwünschten Antwortverhaltens auf die Vertrauensmessung als wenig belastbar, weisen aber auf das Problem der Akquieszenz bei der Beantwortung der Vertrauensitems hin. Ihren Ergebnissen zur Folge hat die Polung der Vertrauensitems erheblichen Einfluss auf das Antwortverhalten der Probanden. So würden positiv formulierte Items im Durchschnitt besser bewertet als negativ formulierte Items. Das heißt: Items, die durch Zustimmung Vertrauen anzeigen, überschätzen den Vertrauenslevel, während Items, die durch Ablehnung Vertrauen anzeigen, den Vertrauenslevel unterschätzen. (Chun/Campbell (1974), S. 1068f; Petermann (2013), S. 24; Rotter (1967), S. 654ff; Wright/Tedeschi (1975), S. 476)

4.2.2 Philosophies of Human Nature (PHN)

Unter *Philosophies of Human Nature* versteht WRIGHTSMAN „attitudes about people in general – attitudes that emphasize the social qualities of people“ (Wrightsmann (1974), S. 28). Sie beschreiben im Wesentlichen die Erwartungen hinsichtlich der Eigenschaften und des Verhaltens von Menschen im Allgemeinen. Insbesondere für die Interaktion mit anderen spielen diese Erwartungen eine große Rolle, da das Handeln anderer oft unerwartet und unvorhersehbar ist. Unsere Reaktionen auf das Verhalten anderer werden dabei maßgeblich durch diese generalisierten Erwartungen bestimmt. Determinanten dieser Erwartungen sind das Verhalten und die Einstellungen der Eltern, der Umwelt, in der man aufwächst, Sozialisationsprogramme, die auf die Entwicklung bestimmter Einstellungen ausgerichtet sind sowie die individuelle Persönlichkeit. (Petermann (2013), S. 22f; Wrightsmann (1964), S. 743; Wrightsmann (1974), S. 28, S. 48)

In diesem Zusammenhang entwickelte WRIGHTSMAN eine Skala, die verschiedene Aspekte dieser Philosophien über die menschliche Natur erfassen sollte. Im Einzelnen waren dies *trustworthiness* (1), *altruism* (2), *independence* (3), *strength of will and rationality* (4), *complexity of Human Nature* (5) und *variability of Human Nature* (6). *Trustworthiness* bezieht sich auf das Ausmaß der Ehrlichkeit, Moralität und Verlässlichkeit anderer. *Altruism* betrifft die Uneigennützigkeit anderer sowie deren aufrichtige Sympathie für und Sorge um andere. *Independence* misst das Ausmaß, in dem man davon überzeugt ist, dass andere ihre Überzeugungen im Angesicht von Konformitätsdruck beibehalten. *Strength of will and rationality* bildet ab, inwieweit man davon überzeugt ist, dass andere ihre eigenen Verhaltensmotive verstehen und inwieweit andere Kontrolle über ihr Leben haben. *Complexity of Human Nature* befasst sich damit, in welchem Ausmaß andere Menschen schwierig bzw. einfach zu verstehen sind, während sich *variability of Human Nature* mit den individuellen Differenzen der menschlichen Natur und der grundlegenden Veränderlichkeit der menschlichen Natur befasst. (Robinson/Shaver/Wrightsmann (1991), S. 385ff; Wrightsmann (1964), S. 744ff; Wrightsmann (1974), S. 41ff)

Die Skala setzte sich in einer frühen Form aus 120 Statements zusammen, die in einem mehrstufigen Prozess von Itemanalysen auf insgesamt 84 Items reduziert werden konnte, wobei jede der 6 Dimensionen der Skala durch 14 Items repräsentiert wurde. Zur Vorbeugung gegen Akquieszenz waren die Items zur einen Hälfte positiv formuliert und zur anderen Hälfte negativ formuliert. Jedes Statement konnte dabei auf einer 6stufigen Ratingskala bewertet werden, deren Ausprägungen mit „*strongly agree*“ (+3), „*moderately agree*“ (+2), „*slightly agree*“ (+1), „*slightly disagree*“ (-1), „*moderately disagree*“ (-2) und „*strongly disagree*“ (-3) benannt wurden. Die Möglichkeit einer neutralen Antwort oder einer *Weiß nicht*-Antwort wurde nicht gegeben. Mit Hilfe von Faktorenanalysen konnten die theoretisch erstellten Dimensionen empirisch nachgewiesen werden. Einschränkend wurde allerdings auch gefunden, dass die Faktoren *trustworthiness* und *altruism* aufgrund ihrer Ähnlichkeit als Teil des gleichen zugrundeliegenden Konstruktes betrachtet werden können. Es zeigte sich außerdem, dass die Dimensionen 1 bis 4 stark miteinander korrelierten, während sie von den Dimensionen 5 und 6 im Wesentlichen unabhängig waren. Während die Dimensionen 1 bis 4 substanzielle Aspekte der menschlichen Natur abfragten, bezogen sich die Dimensionen 5 und 6 auf die Verständlichkeit und die Konsistenz der menschlichen Natur. WRIGHTSMAN nimmt darüber hinaus an, dass die ersten vier Dimensionen einen gemeinsamen Kern besitzen, sieht aber trotzdem die Berechnung eines

4. Messung generalisierten Vertrauens

Summenscores als Index für diese Subskalen kritisch. (Robinson/Shaver/Wrightsman (1991), S. 385ff; Wrightsman (1964), S. 744ff; Wrightsman (1974), S. 49ff, S. 70f)

Eine Analyse der *trustworthiness*-Skala der PHN ergab eine Multidimensionalität anstatt der von WRIGHTSMAN angenommenen Unidimensionalität. So bildet die Vertrauenssubskala auf der einen Seite allgemeine Ehrlichkeit und Moralität, benannt als *global morality*, ab und bezieht sich auf der anderen Seite auf Ehrlichkeit in spezifischen Kontexten (*specific acts of honesty*). Gleichzeitig würden sich 8 der insgesamt 14 Items dieser Subskala als Kurzsкала zur Messung von Vertrauen eignen. (Chun/Campbell (1975), S. 325ff)

4.2.3 Machiavellismus (Mach IV bzw. Mach V)

Die Mach IV bzw. Mach V beziehen sich lediglich indirekt auf Vertrauen, können aber als Vertrauensskalen interpretiert werden. Sie trennen zwischen Personen, die andere Menschen als schwach und manipulierbar ansehen, stark aufgabenorientiert sind und vornehmlich auf eigene Ziele fixiert sind, und Personen, die eher einen persönlichen Umgang pflegen, welche die Perspektive anderer einnehmen können sowie die Ziele und Interessen anderer vor die eigenen Ziele und Interessen setzen. Entsprechend können Personen mit einem hohen Skalenwert als weniger vertrauensvoll und vertrauenswürdig angesehen werden. (Geis (1978), S. 352ff)

Das Konzept des Machiavellismus geht im Wesentlichen auf das von NICCOLO MACHIAVELLI verfasste Werk *Der Fürst* zurück. MACHIAVELLI entwickelte dort Handlungsleitlinien, deren Befolgung den Machterhalt eines Herrschers sicherstellen sollte. So soll der Fürst möglichst öffentlichkeitswirksam die gesellschaftlich verlangten Tugenden befolgen, dabei aber immer vorbereitet sein, jedwede Initiative zu ergreifen, die notwendig ist, um die eigenen Ziele zu erreichen. CHRISTIE nimmt nun an, dass dieses Prinzip für den Menschen an sich Gültigkeit besitzt, sich in seiner Ausprägung allerdings von Individuum zu Individuum unterscheidet. Machiavellismus beschreibt somit eine grundsätzliche Handlungsstrategie für den Umgang mit anderen Menschen, wobei davon ausgegangen wird, dass diese prinzipiell manipulierbar sind und zur Erreichung der eigenen Ziele entsprechend beeinflusst werden sollten. (Christie/Geis (1970), S. 1ff; Geis (1978), S. 305f; Robinson/Shaver/Wrightsman (1991), S. 376ff)

Auf der Basis von *Der Fürst* entwickelte CHRISTIE zunächst eine 71 Statements umfassende Skala (Mach III), die sowohl positiv als auch negativ gepolte Items enthielt. Die Items bildeten dabei insgesamt drei Gegenstandsbereiche ab: (1) *nature of interpersonal tactics* (32 Items), (2) *views of human nature* (28 Items) und (3) *abstract or generalized morality* (11 Items). Der Faktor *nature of interpersonal tactics* umfasste dabei Aussagen zum individuellen Interaktionsverhalten, während der Faktor *views of human nature* Bezug auf die individuellen Überzeugungen hinsichtlich der Natur des Menschen nimmt. *Abstract or generalized morality* betrifft individuelle ethische Überzeugungen. Als Antwortskala wurde eine 6stufige Ratingskala verwendet, deren Ausprägungen mit „agree strongly“ (7), „agree somewhat“ (6), „agree slightly“ (5), „disagree slightly“ (3), „disagree somewhat“ (2) und „disagree strongly“ (1) benannt waren. Wurde keine Antwortalternative ausgewählt wurde dies mit (4) gewertet. Eine „Weiß nicht“-Kategorie existierte nicht. Infolge einer Itemanalyse konnte gezeigt werden, dass 50 der 71 Items ausreichend gut zwischen Personen mit stark ausgeprägten machiavellistischen Tendenzen und Personen mit schwach ausgeprägten machiavellistischen Tendenzen trennen konnten. (Christie/Geis (1970), S. 10ff; Geis (1978), S. 306ff; Robinson/Shaver/Wrightsman (1991), S. 376ff)

4. Messung generalisierten Vertrauens

Im nächsten Schritt wurde die Anzahl der Items reduziert, sodass eine 20 Items umfassende Skala (Mach IV) entstand, die jeweils zur Hälfte aus positiv und negativ formulierten Statements bestand. Dabei betrafen jeweils 9 Items die Faktoren *nature of interpersonal tactics* und *views of human nature*, während 2 Items dem Faktor *abstract or generalized morality* zugerechnet werden konnten. Für die Auswertung wurde ein Basissummenscore von 20 angenommen, sodass eine starke machiavellistische Neigung durch einen Summenscore von 160 angezeigt wurde. Neutralität war bei einem Wert von 100 gegeben, während eine starke Ablehnung machiavellistischer Tendenzen bei einem Wert von 40 angezeigt wurde. Die Mach IV erwies sich dabei allerdings als durch sozial erwünschtes Antwortverhalten beeinflusst, weshalb eine weitere Evolutionsstufe, die Mach V entwickelt wurde. Die Mach V unterschied sich von der Mach IV dadurch, dass anstatt einer 6stufigen Ratingskala nun eine Antwortskala im *Forced-Choice*-Format mit dichotomen Ausprägungen verwendet wurde und dass nun 20 Gruppen mit jeweils drei Items bearbeitet werden sollten. Aufgabe der Probanden war es nun, jeweils das Item auszuwählen, dem sie am meisten zustimmten, und jenes auszuwählen, das sie am meisten ablehnten. Das dritte Item wurde nicht bewertet. Die Itemgruppen enthielten dabei jeweils ein Item, das in Richtung Machiavellismus gepolt war, sowie eines, das in entgegengesetzter Richtung gepolt war. Ergänzt wurden diese beiden durch ein sogenanntes Pufferitem. Die Items unterschieden sich dabei nach ihrer Anfälligkeit für soziale Erwünschtheit. (Christie/Geis (1970), S. 15ff, S. 33f; Geis (1978), S. 307ff; Robinson/Shaver/Wrightsman (1991), S. 376ff)

4.2.4 The Trust Inventory

COUCH/ADAMS/JONES entwickelten mit dem *Trust Inventory* ein Erhebungsinstrument, das simultan drei Vertrauentypen erfassen sollte. Ziel war es, gleichzeitig und abgegrenzt voneinander, Vertrauen in spezifische Beziehungspartner (*partner trust*), Vertrauen in das eigene soziale Netzwerk aus Familie und Freunden (*network trust*) und Vertrauen in Menschen im Allgemeinen bzw. die menschliche Natur (*generalized trust*) zu erfassen. (Couch/Adams/Jones (1996), S. 305, S. 311; Couch/Jones (1997), S. 319f, S. 323)

Zunächst wurde eine 83 Items umfassende Skala erstellt, wobei 32 Items auf *partner trust*, 20 Items auf *network trust* und 31 Items auf *generalized trust* entfielen. In einer Itemanalyse wurde die Anzahl der Items reduziert, sodass die Skala nun insgesamt 50 Items enthält, von denen 20 Items eine *partner trust scale*, 10 Items eine *network trust scale* und 20 Items eine *generalized trust scale* bilden. Es ergibt sich, dass die drei Vertrauensaspekte zwar stark miteinander korrelieren, allerdings trotzdem voneinander zu unterscheidende Konstrukte darstellen. Dieser Umstand kann durch eine Faktorenanalyse zur Konstruktvalidierung bestätigt werden, wonach eine dreifaktorielle Lösung die Variante mit der größten Erklärungskraft darstellt. *Network trust* stellt, wie von COUCH/ADAMS/JONES bereits prognostiziert, einen intermediären Faktor dar, der zwischen *partner trust* und *generalized trust* angesiedelt ist. Entsprechend ergeben sich Überlappungen mit diesen beiden Faktoren. Nicht zuletzt aus diesem Grund besteht die finale Skala des Trust Inventory aus 40 Items, die lediglich Aspekte von *partner trust* und *generalized trust* abfragen. (Couch/Adams/Jones (1996), S. 311ff; Couch/Jones (1997), S. 323ff)

4.2.5 Inventar zur Erfassung des interpersonellen Vertrauens (IIV)

KASSEBAUM entwickelte das Inventar zur Erfassung des interpersonellen Vertrauens aus der Überzeugung heraus, dass insbesondere im deutschsprachigen Raum ein Erhebungsinstrument zur Erfassung verschiedener Aspekte interpersonalen Vertrauens fehlt. Ausgangspunkt der Skalenentwicklung sind Interviews, etablierte Fragebögen und eigene Überlegungen, aus denen sich eine Sammlung von 1860 Items ergibt. Diese Items wurden auf 30 Kategorien aufgeteilt, die das Vertrauen in verschiedene Personen, Personengruppen und Institutionen betrafen. In einem ersten Schritt wurden diejenigen Items eliminiert, die ähnlich formuliert waren oder nicht eindeutig genug auf Vertrauen bezogen waren. Es resultierten 420 Items, die infolge einer Expertenbeurteilung weiter auf 190 Items gekürzt werden konnten. Durch ein weiteres Experten- und Laienrating wurde die Anzahl der Items weiter auf 86 reduziert. Der nächste Schritt umfasste eine teststatistische Analyse der verbleibenden Items, nicht zuletzt um deren Anzahl weiter zu verringern. Letztlich ergab sich eine Skala, die 55 Items umfasste. Diese Skala deckte nun die Dimensionen Vertrauen in Freunde, Partnervertrauen, Allgemeines Vertrauen, Vertrauen in Nachbarn, Vertrauen in Psychotherapeuten und Leichtgläubigkeit ab. (Kassebaum (2004), S. 226ff; Petermann (2013), S. 35f)

4.2.6 KUSIV3

Mit der KUSIV3 haben BEIERLEIN/KEMPER/KOVALEVA/RAMMSTEDT eine ökonomische Skala zur Erfassung *generalisierten Vertrauens* entwickelt, die an der Definition von ROTTER für interpersonales Vertrauen orientiert ist.

„Nach dieser Definition beschreibt das Konstrukt des interpersonalen Vertrauens systematische Persönlichkeitsunterschiede in der individuellen Erwartung, sich auf Worte, Versprechen, mündliche oder schriftliche Äußerungen anderer Personen verlassen zu können.“ (Beierlein/Kemper/Kovaleva/Rammstedt (2012), S. 16)

Die KUSIV3 setzt sich aus insgesamt drei Items zusammen, wobei zwei dieser Items der Skala SOEPtrust entnommen sind. Beide Items weisen im Wortlaut keinen expliziten Bezug zu einer spezifischen Situation oder Personengruppe auf. Das dritte Item wurde aus einem im Rahmen der Skalenentwicklung erstellten Itempool ergänzt und entspricht ebenfalls den zuvor genannten Spezifikationen. Die Bewertung der Items erfolgt dabei anhand einer fünfstufigen Ratingskala mit den folgenden Ausprägungen: „stimme gar nicht zu“ (1), „stimme wenig zu“ (2), „stimme etwas zu“ (3), „stimme ziemlich zu“ (4), „stimme voll und ganz zu“ (5). Das Ausmaß *generalisierten Vertrauens* ergibt sich als Aggregat der individuellen Ausprägungen auf den Einzelitems der Skala. (Beierlein/Kemper/Kovaleva/Rammstedt (2012), S. 8ff)

Die Validität der Skala wurde im Wesentlichen durch die Bestimmung des Zusammenhangs mit aus der Fachliteratur bekannten typischen Korrelaten bestimmt. Hierbei ergaben sich positive Assoziationen mit psychologischen Merkmalen wie Verträglichkeit, Optimismus, Lebens-, Arbeits- und Partnerzufriedenheit sowie der allgemeinen Selbstwirksamkeit. Negative Zusammenhänge ergaben sich mit Pessimismus, Neurotizismus sowie psychischen und physischen Beeinträchtigungen. Mit sozioökonomischen Variablen wie Bildung und Einkommen konnten ebenfalls positive Zusammenhänge gefunden werden. Dagegen standen Alter und Geschlecht nicht mit *generalisiertem Vertrauen* in Verbindung. Die KUSIV3 korrelierte außerdem

positiv mit dem Vertrauen in politische Institutionen, der Größe des sozialen Netzwerkes, der internalen und externalen Political Efficacy sowie dem politischen Engagement. Mit Ausnahme des Zusammenhangs entsprachen alle Assoziationen in Richtung und Stärke den innerhalb früherer Studien postulierten Befunden. (Beierlein/Kemper/Kovaleva/Rammstedt (2012), S. 13f, S. 16f)

4.3 Methodische und inhaltliche Probleme der Standarditems

Das Standardinstrument zur Messung *generalisierten Vertrauens* ist keinesfalls unumstritten, wobei sich Kritik sowohl auf methodische als auch inhaltliche Aspekte beziehen. Ein fundamentaler inhaltlicher Kritikpunkt ergibt sich dabei aus der Negierung der Existenz *generalisierten Vertrauens*, gefolgt vom hinsichtlich der Methodik relevantesten Problem, dem durch die Standarditems abgedeckten Vertrauensradius. Daneben spielen auch die Erfassung der Intensität *generalisierten Vertrauens*, reaktive Effekte, die Multidimensionalität der Standarditems sowie die Differenziertheit der Vertrauensmessung eine Rolle.

4.3.1 Existenz *generalisierten Vertrauens*

Generalisiertes Vertrauen stellt keinesfalls ein allgemein anerkanntes Konzept dar. Kritik entzündet sich dabei vornehmlich daran, dass es sich bei *generalisiertem Vertrauen* um eine unspezifische Form des Vertrauens handelt, dessen Bezugspunkt der generalisierte Andere ist bzw. Menschen im Allgemeinen sind, unabhängig davon, wer diese anderen Menschen sind. Allerdings scheint es unmöglich und wenig plausibel, dass man jedem Menschen auf der Welt im gleichen Maße Vertrauen schenkt bzw. allen Menschen gegenüber vertrauensvoll eingestellt ist. *Generalisiertes Vertrauen* differenziert weder zwischen Personen aus unterschiedlichen Kategorien noch hinsichtlich der Vertrauensobjekte. Es erscheint dagegen viel glaubwürdiger, dass es sich bei Vertrauen um ein erfahrungsgeleitetes und relationales Konzept handelt, sodass sich Vertrauen gegenüber anderen aus vergangenen Erfahrungen mit eben diesen Personen oder vergleichbaren Personen ergibt. Je mehr positive Vertrauenserfahrungen man gemacht hat, auf umso mehr Personen bzw. Personengruppen lässt sich dann in Zukunft Vertrauen übertragen. Gleichzeitig bedeutet ein solches Verständnis von Vertrauen auch nicht, dass man diesen Personen oder Gruppen in jeder Hinsicht vertrauen würde. *Generalisiertes Vertrauen* im Sinne eines grundsätzlichen Vertrauens gegenüber Menschen würde diesen relationalen Charakter des Vertrauens negieren. (Cook/Hardin/Levi (2005), S. 10ff; Hardin (2002), S. 60ff; Hardin (2006), S. 60ff)

In Bezug auf die Standarditems wird nun kritisiert, dass diese den unterschiedlichen Vertrauenskonzeptionen nicht gerecht werden bzw. diese vollkommen außer Acht lassen. Eine Vertrauensmessung mit Hilfe der Standarditems würde dem Befragten die Definition dessen überlassen, was unter Vertrauen zu verstehen ist, bzw. von den Forschenden würde eine allgemeingültige Auffassung von Vertrauen unterstellt, die aber nicht bewiesen ist. Es ist vielmehr anzunehmen, dass sich das Verständnis von Vertrauen zwischen einzelnen Personen oder Personengruppen mehr oder weniger stark unterscheidet. Die Konsequenz ist, dass Daten, die mit Hilfe der Standarditems gewonnen wurden, nicht eindeutig interpretiert werden

4. Messung generalisierten Vertrauens

können, da unklar bleibt, welches die jeweiligen Bezugspunkte des auf diese Weise gemessenen Vertrauens sind. (Cook/Hardin/Levi (2005), S. 10ff; Hardin (2002), S. 60ff; Hardin (2006), S. 60ff)

Problematisch am Verständnis von Vertrauen als rein relationales Konzept ist der Umstand, dass solches Vertrauen eine Interaktionshistorie zwischen den Vertrauenspartnern voraussetzt. Grundvoraussetzung für jegliche Interaktion bzw. Kooperation ist allerdings ein Mindestmaß an Vertrauen.

„Consider two people who will join us in this chapter: Bill and Jane. If Jane trusts Bill to keep his word and if Bill trusts Jane to keep her word, they can reach an agreement to cooperate and thus make each other better off. Even without some external enforcement mechanism (such as an arbitrator, the police, or the courts), they will keep to their agreements.

If Jane and Bill did not know each other, they would have no basis for trusting each other. Moreover, a single encounter will not suffice to develop trust. Jane and Bill have to interact over time to develop reputations for keeping their word. And, even when they get to know each other better, their mutual trust will be limited to what they know about each other.“ (Uslaner (2002), S. 16)

Wie in diesem Beispiel sichtbar wird, hängt die Bereitschaft von Jane und Bill miteinander zu kooperieren im Wesentlichen davon ab, ob sie sich wechselseitig vertrauen können. Ein relationales Vertrauen würde zu seiner Etablierung Informationen über den Vertrauenspartner und dessen Einstellung gegenüber dem Vertrauensgegenstand benötigen. Im Idealfall wäre die Situation, aus der sich diese Informationen generieren, äquivalent zur aktuellen Vertrauenssituation. Allerdings würde eine einzelne Beobachtung des Verhaltens des Vertrauenspartners noch keine hinreichende Evidenz für seine Vertrauenswürdigkeit darstellen, hierzu wären mehrfache Beobachtungen notwendig. Hätte Vertrauen einen solchen grundlegend relationalen Charakter, würde es unmöglich neue Beziehungen zu bisher unbekannten Personen aufzubauen bzw. die Interaktionen des täglichen Lebens zu bestreiten, in denen man in hohem Maße gezwungen ist, mit Fremden umzugehen oder in denen man auf Leute trifft, denen man zum ersten Mal begegnet. Gesellschaft, deren Existenz vom Bestand gewisser Regeln bzw. einer sozialen Ordnung abhängig ist, auf deren Allgemeingültigkeit vertraut wird, würde ebenso unmöglich. *Generalisiertes Vertrauen* reduziert in beiden Fällen die Komplexität der sozialen Situation und macht auf diese Weise eine Interaktion zwischen den Gesellschaftsmitgliedern möglich. Dies geschieht auf der einen Seite dadurch, dass der Vertrauenspartner als generalisierter Anderer wahrgenommen wird und auf der anderen Seite dadurch, dass die wechselseitige Überzeugung vorherrscht, dass man den gleichen Verhaltensnormen folgt und unterworfen ist. (Barber (1983), S. 10ff, S. 19ff; Garfinkel (1963), S. 190, S. 193f, S. 238; Garfinkel (1967), S. 35, S. 173; Uslaner (2002), S. 21)

Negiert man die Existenz eines *generalisierten Vertrauens* und geht man gleichzeitig von einem grundlegend relationalen Vertrauen aus, so würde man sich somit einem Henne-Ei-Problem gegenübersehen. Relationales Vertrauen würde Informationen aus einer vorangegangenen Interaktion zur Generierung von Vertrauen voraussetzen, während eine solche Interaktion bereits ein gewisses Maß an Vertrauen voraussetzt. Dieses Dilemma kann nur durch die Annahme einer generalisierten Vertrauenseinstellung überwunden werden.

Neben dieser theoretischen Herleitung der Notwendigkeit eines *generalisierten Vertrauens* existieren auch empirische Belege für dessen Existenz. So wurde im Rahmen der ANES (American National Election Study) im Jahr 2000 ein „*think aloud*“-Experiment durchgeführt, bei dem Teile der Befragten im Sinne eines kognitiven Pretests angeben sollten, welche Bedeutung die verschiedenen Standarditems jeweils für sie hatten. Bezüglich der Vertrauensdimension der

4. Messung generalisierten Vertrauens

Standarditems gingen dabei mehr als 70% der Befragten davon aus, dass dort ein allgemeines Vertrauen abgefragt wird, welches losgelöst von Erfahrung existiert. Bezüglich der Fairnessdimension gingen über 50% der Befragten davon aus, dass es sich hier um die Abfrage einer generellen Fairnessüberzeugung handelt. Bei der Hilfsbereitschaftsdimension gingen immerhin noch knapp 40% der Befragten davon aus, dass mit Hilfe des entsprechenden Items eine grundsätzliche Hilfsbereitschaftsüberzeugung abgefragt werden sollte. (Uslaner (2002), S. 72ff; Uslaner (2012b), S. 75; Uslaner (2012c) S. 103)

STURGIS/SMITH führten eine Replikation des „*think aloud*“-Experiments durch und fanden, dass 35% der Befragten bei der Beantwortung der Standarditems an ihnen unbekannte Personen gedachten hatten, allerdings hatten auch 28% an ihnen bekannte Personen gedacht. (Sturgis/Smith (2010), S. 87)

Weiterer empirischer Beleg für die Existenz *generalisierten Vertrauens* ist die Beständigkeit der mit Hilfe der Standarditems ermittelten Vertrauenswerte. Würde es sich bei dem mit den Standarditems erfassten Vertrauen um ein relationales Vertrauen handeln, müssten sich starke Schwankungen ergeben. Die Stabilität der Messwerte spricht allerdings dafür, dass es sich eher um eine moralisch und normativ begründete Form eines Vertrauens mit grundsätzlichem Charakter handelt. Als Beleg für die angesprochene Stabilität der Vertrauenswerte führt USLANER die Ergebnisse des ANES Panel über die Jahre 1972, 1974 und 1976 sowie für die Jahre 2004 bis 2006 an. (Uslaner (2002), S. 57ff; Uslaner (2012b), S. 75f; Uslaner (2012c) S. 106)

4.3.2 Radius *generalisierten Vertrauens*

Das Radiusproblem stellt eine der gewichtigsten Herausforderungen der Messung *generalisierten Vertrauens* dar. Als Radius des Vertrauens wird dabei der Zirkel an Personen beschrieben, innerhalb dessen man bereit ist, Vertrauens- bzw. Kooperationsbeziehungen einzugehen. Im Allgemeinen richtet sich der Vertrauensradius danach, ob man Personen außerhalb der eigenen *ingroup* positiv oder negativ gegenübersteht. Eine stark ausgeprägte *ingroup*-Solidarität reduziert dabei gleichzeitig die Bereitschaft mit Außenseitern zu kooperieren sowie das Vertrauen, das diesen Personen entgegengebracht wird. Allerdings ist der Vertrauensradius auch nicht automatisch auf die gesamte *ingroup* ausgedehnt. So kann sich Vertrauen auch innerhalb der *ingroup* nur auf einen bestimmten Kreis an Personen oder sogar nur einzelne Individuen beziehen. (Delhey/Newton/Welzel (2011), S. 787ff; Fukuyama (2000), S. 4f; Fukuyama (2001), S. 8f)

Die methodische Diskussion des Radiusproblems beschäftigt sich damit, auf welche Art und Weise Befragte die Phrase „*most people*“ innerhalb der Standarditems interpretieren. Vordergründig scheint klar, dass die gewählte Formulierung auf einen generalisierten Anderen als Bezugspunkt des abgefragten Vertrauens hinweist. Bei näherer Betrachtung fällt allerdings auf, dass es höchst fragwürdig ist, ob „*most people*“ für alle Befragten die gleiche Bedeutung hat. (Bjørnskov (2006), S. 2; Delhey/Newton/Welzel (2011), S. 787ff; Glaeser/Laibson/Scheinkam/Soutter (2000), S. 815; Reeskens/Hooghe (2008), S. 516) HARDIN²⁴ fasst diesen Umstand recht drastisch zusammen:

²⁴ Nach HARDIN handelt es sich bei Vertrauen um eine dreiwertige Relation, die die Form „A trusts B to do X“ (Hardin (2002), S. 9) annimmt. Vertrauen wird seiner Auffassung nach somit immer einem konkreten Anderen bezüglich eines bestimmten Objektes entgegengebracht. *Generalisiertes Vertrauen* stellt nach HARDIN lediglich eine einwertige oder allenfalls eine zweiwertige Relation dar. (Hardin (2002), S. 60ff) Vertrauen als einwertige Relation bedeutet „I trust out of a pure disposition to trust anyone and everyone with respect to anything and everything, in which case I am

4. Messung generalisierten Vertrauens

„Hence, either the surveyors ask for insane responses, or respondents to these surveys have to create their own universe of people and objects to which they then apply the questions. You might narrow the population down from everyone to merely those with whom you deal, that is, to the category that would fit trust as a relational concept. Others might narrow it to those whom they might encounter on the sidewalks and in the shops of their city. Perhaps people narrow the field today in relatively different ways from how they narrowed it several decades ago. Who can possibly know what the response to such ill-defined questions tell us about actual levels of trust or about trends in trust over time in any well-defined context?“ (Hardin (2006), S. 61f)

An anderer Stelle heißt es außerdem:

„If I say I can trust most people most of the time, I merely be saying I trust most of those I deal with most of the time. Of course, that is partly why I deal with them and not lots of other people whom I would not trust most of the time.“ (Hardin (2002), S. 61)

Dafür, dass von den Befragten bei der Beantwortung der Standarditems eher ein Vertrauen in konkrete Andere angegeben wird als ein Vertrauen gegenüber generalisierten Anderen, spricht außerdem, dass das Vertrauen zu Menschen abnimmt, je weiter außerhalb sich diese vom Zirkel der vertrauten Personen, wie Familie, Freunde, Kollegen, Nachbarn oder Bekannte, befinden. Die objektive Unbestimmtheit der Bezugsgröße des gemessenen Vertrauens stellt die Validität der Standarditems somit gleich auf zwei Arten in Frage. Erstens kann nicht bestimmt werden, ob *partikulares* oder *generalisiertes Vertrauen* erfasst wird, da unklar ist, wen der Befragte im Moment der Bearbeitung der Standarditems im Kopf hat. Zweitens ist eine Vergleichbarkeit der Vertrauens-Scores einzelner Individuen nicht möglich, da nicht sichergestellt werden kann, ob für beide Vertrauens-Scores der gleiche Bezugsrahmen gilt. Dies hat auch Auswirkungen auf die Makroebene, sodass auch keine Vergleiche mehr zwischen Gruppen, Ethnien oder Nationen möglich wären. (Delhey/Newton/Welzel (2011), S. 787ff; Glaeser/Laibson/Scheinkam/Soutter (2000), S. 815; Reeskens/Hooghe (2008), S. 516)

HARDIN kritisiert zusätzlich zur Verwendung der Formulierung „*most people*“ auch den Einsatz der Phrase „*most of the time*“ innerhalb der Standarditems und erweitert das Radiusproblem auf diese Weise um einen weiteren Aspekt:

„The standard survey question refers to ‚most people‘; but then, it also says ‚most of the time‘, which could be taken to mean that the respondents must restrict the range of matters on which they would trust ‚most people‘.“ (Hardin (2002), S. 61)

„They do not differentiate varied categories of people whom one would be more or less likely to trust, and they do not differentiate different objects of trust ranging from reciprocating minor favors to fulfilling major, very costly promises. The questions generally ask about trusting everyone or most people, implicitly with respect to anything at all. In the formula A trust B with respect to X both B and X in the survey instruments often implicitly roam over ranges of everyone and everything. No sane person trusts everyone equally and with respect to any and every level of risk.“ (Hardin (2006), S. 61)

Innerhalb der Standarditems bleibt somit neben der Bestimmung des Vertrauenspartners auch die Eingrenzung des Vertrauensobjektes dem jeweils Befragten überlassen. Es ergeben sich so

the only variable part“ (Hardin (2002), S. 9), während Vertrauen als zweiwertige Relation die Form „A trusts B, without any conditional constraint on the scope of the trust“ (Hardin (2002), S. 60) annimmt. Die Kritik HARDINS an den Standarditems zur Messung *generalisierten Vertrauens* ist immer vor dem Hintergrund dieser Unterscheidung und seiner Ablehnung des Konzeptes „*generalisierten Vertrauens*“ zu betrachten.

4. Messung generalisierten Vertrauens

auf der Ebene des Vertrauensobjektes die gleichen messtheoretischen Probleme wie schon zuvor auf der Ebene des Vertrauenspartners. Es bleibt unbestimmt, welches Vertrauensobjekt als Basis für den Vertrauens-Score herangezogen wurde. Je nachdem, ob es sich bei dem Vertrauensobjekt um einen Gegenstand handelt, der einen besonderen Wert für den Befragten hat, spielt auch der gewählte Vertrauenspartner als Bezugsgröße eine große Rolle. Mitunter wird unterschiedlichen Personen hinsichtlich verschiedener Objekte unterschiedlich stark vertraut. Entsprechend kann davon ausgegangen werden, dass die beiden hier diskutierten Aspekte des Radiusproblems miteinander korrespondieren und sich wechselseitig beeinflussen. Außerdem ist auch hier wieder kein objektiver Vergleich zwischen verschiedenen Befragten möglich, da sich die jeweiligen Bezugsrahmen des Vertrauens mehr oder weniger stark voneinander unterscheiden könnten.

Innerhalb der Literatur existieren verschiedene Lösungen für das ursprüngliche Radiusproblem, die allerdings einem gemeinsamen Ansatz folgen. Im Wesentlichen werden Items formuliert, die jeweils Vertrauen in eine spezifische Personengruppe abfragen, wobei sich lediglich die Anzahl der Ausdifferenzierungen der Gruppen von Instrument zu Instrument unterscheiden.

WELZEL unterscheidet innerhalb seiner Skala in *out-group trust* und *in-group trust*. Unter ersteres fällt dabei Vertrauen gegenüber Personen, die man zum ersten Mal getroffen hat sowie Personen die anderen Religionsgemeinschaften oder Nationalitäten angehören. *In-group trust* betrifft dagegen Vertrauen gegenüber der Familie, Nachbarn und Personen, die einem persönlich bekannt sind. Es ergibt sich auf diese Weise folgender Wortlaut für das Item (Delhey/Newton/Welzel (2011), S. 790ff; Welzel (2010), S. 162):

Abb. 4.6: Itembatterie zur Differenzierung von *in-group* und *out-group trust* nach WELZEL

„I'd like to ask you how much you trust people from various groups. Could you tell me for each whether you trust people from this group completely, somewhat, not very much or not at all?“

„Your family“ „Your neighborhood“ „People you know personally“	<i>in-group trust</i>
„People you met for the first time“ „People of another religion“ „People of another nationality“	<i>out-group trust</i>

(Quelle: Delhey/Newton/Welzel (2011), S. 792)

NAEF/SCHUPP konzentrieren sich bei ihrer Lösung für das Radiusproblem auf die Abfrage von Vertrauen gegenüber Fremden und formulieren die folgenden Items, die jeweils auf einer 4stufigen Ratingskala mit je nach Statement unterschiedlichen Ausprägungen beantwortet werden konnten (Naef/Schupp (2009), S. 7):

Abb. 4.7: Differenzierung von *trust* und *caution* nach NAEF/SCHUPP

Trust (Vertrauen)	„How much do you trust strangers you meet for the first time.“			
	„no trust at all“	„little trust“	quite a bit of trust“	„a lot of trust“
Caution (Vorsicht)	„When dealing with strangers, it's better to be cautious before trusting them.“			
	„Disagree strongly“	„Disagree somewhat“	„Agree somewhat“	„Agree strongly“

(Quelle: Naef/Schupp (2009), S. 7)

WELCH/SIKKINK/LOVELAND differenzieren zwischen einem Vertrauen in Bekannte, die sie von Familie und Freunden abgrenzen, und einem Vertrauen gegenüber Fremden. Bewertet wurden beide Items auf einer 4stufigen Ratingskala, deren Ausprägungen „*not at all*“, „*only a little*“, „*some*“ und „*a lot*“ lauteten (Welch/Sikkink/Loveland (2007), S. 27f):

„Outside of your family and friends, how much, if at all, do you generally trust people you know personally?“

„How about people you do not know personally? Do you generally trust them...?“

COUCH ET AL. unterstreichen mit ihrer Arbeit noch einmal, dass tatsächlich ein Unterschied zwischen dem Vertrauen in konkrete Andere und generalisierte Andere besteht und dass hierbei die Differenzierung in bekannte und fremde Personen eine große Rolle spielt. Ziel von COUCH ET AL. war es ein Messinstrument zu entwickeln, das eine Simultanmessung von drei verschiedenen Vertrauensformen ermöglicht. Es handelte sich hierbei um *generalized trust* (allgemeines Vertrauen in die menschliche Natur), *partner trust* (Vertrauen in spezifische Beziehungen zu bestimmten Personen) und einem zwischen diesen beiden Vertrauensformen angesiedelten *network trust* (Vertrauen in soziales Netzwerk aus Familie und Freunden). Eine Analyse des sogenannten *trust inventory* bestätigte die konzeptionelle Unterscheidung zwischen *generalized trust* und *partner trust*. So korrelierten Items der *generalized trust scale* eher mit externen Maßen für ein globales Vertrauen und Items der *partner trust scale* korrelierten eher mit externen Maßen für relationales Vertrauen, aber nicht umgekehrt. Die *network trust scale* korrelierte dagegen signifikant, aber moderat mit sowohl relationalem als auch globalem Vertrauen. (Couch/Adams/Jones (1996), S. 318f; Couch/Jones (1997), S. 320, S. 331ff)

Auch wenn auf Basis dieser Untersuchung noch nicht festgestellt werden konnte, ob es sich bei *network trust* in gleicher Art und Weise um eine disjunkte Vertrauenskategorie handelt wie bei *generalized trust* und *partner trust*, so sind die Ergebnisse ein starker Beleg für die Existenz des Radiusproblems. So scheint die Differenzierung in verschiedene Bezugsgruppen des Vertrauens einen Effekt auf das erwiesene Vertrauen zu haben, gleichzeitig ist aber fragwürdig, ob es sich bei dem in dieser Art und Weise differenziert gemessenen Vertrauen noch um *generalisiertes Vertrauen* handelt.

Einen Beleg dafür, dass es sich bei dem mit den Standarditems erfassten Vertrauen in den meisten Fällen doch um eine generalisierte Vertrauenseinstellung handelt, findet sich bei STURGIS/SMITH. Ihre Analyse der Standarditems im Kontrast zu einem Vertrauensitem, das speziell auf das Vertrauen in Personen aus dem näheren Umfeld der Befragten ausgerichtet war, sollte im Wesentlichen die folgenden Hypothesen prüfen (Sturgis/Smith (2010), S. 77ff):

4. Messung generalisierten Vertrauens

„H1: When answering the GTQ²⁵, respondents who report thinking about people with whom they are personally acquainted will have a higher probability of reporting trust than respondents who draw on more generalized categories, such as ‘people in general’.“

„H2: The proportion of respondents reporting that they think of people with whom they are personally acquainted will be significantly higher for the TiN²⁶ item than for the GTQ.“

„H3: Holding constant the response format of the questions, the TiN item will produce substantially higher estimates for trust than the GTQ.“

Innerhalb ihrer Analyse wurden die Befragten zunächst per Zufall in zwei Experimentalgruppen verteilt. Die erste Gruppe bekam die GTQ präsentiert, während die zweite Gruppe das TiN-Item bearbeiten sollte. An das jeweilige Vertrauensitem angeschlossen wurde außerdem die folgende Frage gestellt:

„In answering the last question, who came to mind when you were thinking about ‘most people’/‘people in your local area’?“

Dabei wurden die Befragten vom Interviewer solange nach weiteren Personen gefragt, bis diese keine zusätzlichen mehr nannten. (Sturgis/Smith (2010), S. 80)

Die Analyse ergab, dass Befragte, die bei der Beantwortung der GTQ bekannte Personen als Referenz nutzen, ein höheres Vertrauen aufwiesen als Befragte, die an Menschen im Allgemeinen gedacht hatten. Außerdem ergab sich, dass Befragte, die die GTQ bearbeitet hatten, verglichen mit Befragten, die das TiN bearbeitet hatten, eine um 50% reduzierte Wahrscheinlichkeit besaßen, bekannte Personen als Referenz für ihr Vertrauensurteil zu verwenden. Gleichzeitig war die Wahrscheinlichkeit Menschen im Allgemeinen als Referenz zu benutzen für die Befragten der TiN-Gruppe viermal niedriger als in der GTQ-Gruppe. Zusätzlich ist die Wahrscheinlichkeit Vertrauen anzuzeigen um 19% höher, wenn das Vertrauensitem auf Personen aus dem näheren Umfeld beschränkt ist. Die TiN-Gruppe wies somit höhere Vertrauenswerte auf als die GTQ-Gruppe. (Sturgis/Smith (2010), S. 83, S. 87f)

STURGIS/SMITH schlussfolgern, dass eine substantiell größere Anzahl an Befragten (34%) bei der Bearbeitung der GTQ eher Fremde als Referenz nutzen als persönlich bekannte Personen (25%). Dies deutet daraufhin, dass die GTQ tatsächlich *generalisiertes Vertrauen* misst.

4.3.3 Vertrauenslevel

Während der Vertrauensradius den Geltungsbereich des Vertrauens festlegt, beschreibt der Vertrauenslevel die Intensität des Vertrauens das anderen Personen bzw. den Personen

²⁵ Rosenberg Generalized Trust Question (GTQ):

„Generally speaking would you say that most people can be trusted, or that you can’t be too careful in dealing with people?

- a. Most people can be trusted
- b. Can’t be too careful“

²⁶ Trust in Neighbors (TiN):

„How much do you trust people in your local area?

- a. A lot
- b. A fair amount
- c. Not very much
- d. Not at all“

4. Messung generalisierten Vertrauens

innerhalb dieses Zirkels entgegengebracht wird. Der Vertrauenslevel determiniert dabei die Bereitschaft des Einzelnen mit anderen zu kooperieren. Zunächst stellen Vertrauensradius und Vertrauenslevel zwei voneinander streng zu unterscheidende Konzepte dar, auf den zweiten Blick ergibt sich allerdings trotzdem eine Verbindung, die insbesondere für die Messung von Vertrauen eine besondere Tragweite besitzt. So ist ein hoher Vertrauenslevel nicht an einen weiten Vertrauensradius gebunden oder umgekehrt. Vielmehr können sowohl ein hoher als auch ein niedriger Vertrauenslevel auf einen engen oder weiten Vertrauensradius bezogen sein. (Delhey/Newton/Welzel (2011), S. 787f; Delhey/Newton/Welzel (2014), S. 1260f; van Hoorn (2014), S. 1256; van Hoorn (2015), S. 269f) Anders ausgedrückt: Vertrauenslevel und Vertrauensradius können sich nicht wechselseitig ausgleichen, sodass ein hoher Vertrauenslevel einen engen Vertrauensradius kompensieren bzw. ein weiter Vertrauensradius einen niedrigen Vertrauensradius kompensieren würde. Im ersten Fall würde die betrachtete Person schlichtweg über ein hohes *partikulares Vertrauen* verfügen, während die betrachtete Person im zweiten Fall über ein niedriges *generalisiertes Vertrauen* verfügen würde. Entsprechend könnten sich außerdem die Paarungen niedriger Vertrauenslevel und enger Vertrauensradius sowie hoher Vertrauenslevel und weiter Vertrauensradius ergeben. Die erste Paarung bedeutet dann ein niedriges *partikulares Vertrauen*, während die zweite Paarung ein hohes *generalisiertes Vertrauen* anzeigt.

„The level and radius of trust do not compensate each other, they condition each other: a high level of trust produces a large amount of general trust only if the radius of trust is wide; a wide radius of trust produces a large amount of general trust only if the level is high.“ (Delhey/Newton/Welzel (2011), S. 788)

Gleichzeitig wird aber in keiner der hier aufgezeigten Kombinationen eine Aussage über das Ausmaß des jeweils entgegengesetzten Vertrauens gemacht. Erhält man eine Information über die Intensität des *generalisierten Vertrauens*, bedeutet dies noch keine Information über die Intensität des *partikularen Vertrauens*. (van Hoorn (2014), S. 1258)

Die Problematik der Bestimmung des Vertrauenslevels tritt insbesondere im Hinblick auf die Erfassung *generalisierten Vertrauens* mit Hilfe der Standarditems auf. So wurde weiter oben im Zusammenhang mit dem Radiusproblem dargestellt, dass durch die Phrase „*most people*“ innerhalb der Standarditems nicht spezifiziert wird, in Bezug auf welchen Personenkreis der individuelle Vertrauenslevel angegeben werden soll. Die Definition des Bezugsrahmens bleibt vollkommen dem jeweiligen Befragten überlassen. STURGIS/SMITH fanden unterdies heraus, dass bei der Beantwortung der Standarditems zwar eher fremde Personen als Referenz verwendet werden, aber gleichzeitig eine nicht zu verachtende Anzahl der Befragten ausschließlich ihnen bekannte Personen als Grundlage ihrer Vertrauensbestimmung nutzen. Diese Vermischung der Messung von *generalisiertem* und *partikularem Vertrauen* kann dabei ernsthafte Auswirkungen auf die Validität der ermittelten Vertrauenswerte haben. So ergibt sich in der Regel ein höheres Vertrauen, wenn bekannte Personen als Bezugspunkt herangezogen werden, als wenn generalisierte Andere als Referenz verwendet werden. Dies kann schlussendlich zu einer Überschätzung des Ausmaßes an *generalisiertem Vertrauen* führen. (Sturgis/Smith (2010), S. 89f)

4.3.4 Reaktive Effekte

Der Einfluss reaktiver Effekte auf die Messung *generalisierten Vertrauens* wurde bisher nur wenig untersucht und ist dementsprechend selten dokumentiert.

SMITH überprüfte in diesem Zusammenhang, ob die Fragenreihung innerhalb eines Fragebogens Auswirkungen auf die Beantwortung der Standarditems hat. Hierzu wurden den Standarditems in zwei Versuchsreihen unterschiedliche Fragen vorangestellt. Unter der Bedingung *trust1* wurden Abfragen hinsichtlich des Arbeitswertes bzw. der Anforderungen der Arbeit, dem Fortkommen im Leben, der Sexualmoral und dem Umstand, ob das eigene Leben spannend ist, gemacht, bevor die *trust*-Komponente der Standarditems bearbeitet werden sollte. Unter der Bedingung *trust2* wurden hingegen Abfragen bezüglich Kriminalität, politischer Ideologie, Egalisierung von Reichtum, Scheidungsrecht und Legalisierung von Marihuana getätigt, bevor die *trust*-Komponente der Standarditems beantwortet werden sollte. Unter der Bedingung *helpful1* wurden der *helpfulness*-Komponente der Standarditems Fragen zur persönlichen finanziellen Situation, der subjektiven Schichteinstufung, Anomie und Gebrauch von Gewalt vorangestellt. Unter der Bedingung *helpful2* waren es dagegen Fragen zur Mediennutzung, zum Rauchen, Trinken und zur Geselligkeit. Tatsächlich fand SMITH für die beiden getesteten Dimensionen der Standarditems starke Kontexteffekte. So wurde unter *trust2* verglichen mit *trust1* weniger Vertrauen ausgedrückt, was SMITH auf den Kriminalitäts- und Betrugskontext der vorangestellten Fragen unter *trust2* zurückführt. Außerdem wurde unter der Bedingung *helpful1* weniger Hilfsbereitschaft angegeben als unter der Bedingung *helpful2*, was SMITH auf den Anomie- und Gewaltkontext der vorangestellten Items unter der Bedingung *helpful1* zurückführt. Allgemein erklärt SMITH die Existenz von Fragereiheneffekten bei der Messung von *generalisiertem Vertrauen* damit, dass hier nach einer generalisierenden Bewertung von Menschen im Allgemeinen gefragt wurde, für die das jeweilige Individuum seine gesamte Lebenserfahrung Revue passieren lassen muss. Die Beurteilung auf Basis einer solchen Fülle an Informationen ist von Haus aus schon schwierig und wird nochmals erschwert, da es sich um Erinnerungen handelt, die Verzerrungen aufweisen können. (Smith (1997), S. 174ff; Uslaner (2007), S. 3; Uslaner (2012b), S. 74f)

NAEF/SCHUPP untersuchen ebenfalls, ob sich die Positionierung der Standarditems innerhalb eines Fragebogens auf das Antwortverhalten auswirken, konzentrieren sich dabei aber auf mögliche *primacy*- bzw. *recency*-Effekte. Hierzu wurden die Versuchsteilnehmer per Zufall auf zwei Gruppen aufgeteilt, wobei in einer Gruppe die Vertrauensitems am Anfang des Fragebogens gestellt wurden, während die zweite Gruppe die Vertrauensitems erst am Ende des Fragebogens präsentiert bekam. Ein Vergleich des Antwortverhaltens zwischen den Gruppen zeigte, dass die Befragten, denen die Vertrauensitems am Ende des Fragebogens präsentiert wurden, mehr Vertrauen in Fremde besaßen als die Befragten, die die entsprechenden Items am Anfang des Fragebogens beantwortet hatten. Offensichtlich scheint somit die Position der Vertrauensitems innerhalb eines Fragebogens einen Einfluss auf das ermittelte Ausmaß an Vertrauen zu haben. NAEF/SCHUPP schlussfolgern, dass für eine bessere Vergleichbarkeit der Vertrauenswerte zwischen einzelnen Befragten, die Vertrauensitems immer an der gleichen Stelle innerhalb eines Fragebogens positioniert werden sollten. (Naef/Schupp (2009), S. 9f)

Eine weitere Verzerrung der Vertrauenswerte kann sich dadurch ergeben, dass die Vertrauensitems nicht ehrlich beantwortet werden. (Glaeser/Laibson/Scheinkam/Soutter (2000), S. 815) Einer der Hintergründe für ein solches Verhalten seitens der Befragten könnte die folgende Überlegung sein:

„(...), the tendency has been to view ‘trust’ as positive and advantageous, but to see ‘distrust’ as negative and disadvantageous.“ (Lewicki/Brinsfield (2012))

Tatsächlich scheint die Messung von *generalisiertem Vertrauen* durch sozial-erwünschtes Antwortverhalten beeinflusst zu sein. NAEF/SCHUPP untersuchen in diesem Zusammenhang, inwiefern die Standarditems mit den zwei Dimensionen des sogenannten *Balanced Inventory of Desirable Responding*, einer Skala zur Erfassung sozialer Erwünschtheit, zusammenhängen. Die erste Dimension dieser Skala misst die Tendenz der Befragten, die Realität optimistischer wahrzunehmen als angebracht wäre, und wird als *self-deceptive enhancement* („selbsttäuschende Verstärkung“) bezeichnet. Die zweite Dimension betrifft das Ausmaß, in dem eine Person wissentlich ein positives Bild von anderen zu konstruieren versucht, und wird als *impression management* („Kontrolle des Eindrucks“) bezeichnet. Es ergaben sich für beide Dimensionen signifikante Zusammenhänge mit den Standarditems. Dabei überbewerteten Personen, die zu sozialerwünschtem Antwortverhalten neigten, ihr Vertrauen gegenüber Fremden, was eine starke Beeinträchtigung der Validität der Standarditems bedeutet. (Naef/Schupp (2009), S. 10)

4.3.5 Multidimensionalität der Standarditems

Ein weiteres Problem der Vertrauensmessung mit den Standarditems stellt die Multidimensionalität bzw. die sogenannte *double-barreledness* der Vertrauenskomponente der Standarditems dar. Betrachtet man die Vertrauenskomponente, so fällt auf, dass zwei voneinander zu unterscheidende Einstellungen abgefragt werden bzw. die Antwortskala einen bipolaren Charakter hat, wobei sich zwei inhaltlich verschiedene Konstrukte gegenüberstehen. So wird innerhalb des Items nicht zwischen Vertrauen und Misstrauen unterschieden, sondern zwischen den voneinander verschiedenen Konzepten Vertrauen und Vorsicht. Die Vertrauenskomponente fragt demzufolge zwei Konstrukte innerhalb einer Frage ab, sodass im Nachhinein nicht mehr differenziert werden kann, welche dieser beiden Dimensionen für die jeweilige Antwort der Befragten verantwortlich ist. So könnte eine Person beispielsweise ein hohes Maß an *generalisiertem Vertrauen* erreichen, aber gleichzeitig eine starke Risikoaversion besitzen und deshalb zur Vorsicht tendieren. (Miller/Mitamura (2003), S. 63f; Naef/Schupp (2009), S. 6f; Uslaner (2007), S. 4; Uslaner (2012b), S. 75; Uslaner (2012c), S. 102; Yamagishi/Kikuchi/Kosugi (1999), S. 147)

NAEF/SCHUPP versuchen dieses Problem zu lösen, indem sie das Item in zwei getrennte Fragen aufspalten, die jeweils auf einer Ratingskala mit den Antwortoptionen „*Stimme voll zu*“, „*Stimme eher zu*“, „*Lehne eher ab*“ und „*Lehne voll ab*“ bewertet werden sollen:

„Im Allgemeinen kann man den Menschen vertrauen.“

„Heutzutage kann man sich auf niemanden mehr verlassen.“

4.3.6 Differenziertheit der Vertrauensmessung

Die Differenziertheit der Messung *generalisierten Vertrauens* betrifft auf der einen Seite die Anzahl der Items, die zur Erfassung *generalisierten Vertrauens* verwendet werden und auf der anderen Seite die Breite der jeweils angelegten Antwortskala.

Wie bereits beschrieben, wird je nach Bevölkerungsumfrage eine unterschiedliche Anzahl an Vertrauensitems für die Erfassung von *generalisiertem Vertrauen* eingesetzt. So erfasst beispielsweise der WVS *generalisiertes Vertrauen* anhand von einem Item, während der GSS und der ESS drei Items einsetzen. Die Messung *generalisierten Vertrauens* mit nur einem einzelnen Item wird dabei aus verschiedenen Gründen als problematisch angesehen und entsprechend kritisiert. Der erste Kritikpunkt betrifft die Reliabilität des gemessenen Vertrauenslevels. So wird angenommen, dass die Befragten bei der Verwendung von nur einem Item kein über die Zeit konsistentes Antwortverhalten zeigen. Durch den Einsatz multipler Indikatoren könnte die Variation der Antwortmuster über die Zeit dagegen reduziert werden. Der zweite Kritikpunkt betrifft den Mangel an Präzision, der mit der Vertrauensmessung mittels lediglich eines Items einhergeht. Um die kognitive Verarbeitungsfähigkeit der Befragten nicht überzustrapazieren, wird oft die Anzahl der Antwortkategorien begrenzt, was allerdings die Erklärungskraft des Items reduziert. Mittels multipler Indikatoren können dagegen Messskalen gebildet werden, die verschiedene Aspekte *generalisierten Vertrauens* abfragen und auf diese Weise ein präziseres Abbild des Vertrauenslevels der Befragten ermöglichen. Der dritte Kritikpunkt betrifft die mangelhafte Reichweite der Vertrauensmessung mit einem einzelnen Item. Die konzeptuelle und theoretische Komplexität *generalisierten Vertrauens* kann nur mit Hilfe multipler Indikatoren adäquat abgebildet werden. Insgesamt scheint die Verwendung multipler Indikatoren bzw. von Messskalen für die Erfassung *generalisierten Vertrauens* die Anfälligkeit für Messfehler zu verringern und die Reliabilität der Messungen zu verbessern. (Miller/Mitamura (2003), S. 69; Reeskens/Hooghe (2008), S. 519f, S. 530; Uslaner (2012b), S. 76)

Neben der unterschiedlichen Anzahl an Items, die je nach Umfrage zur Erfassung *generalisierten Vertrauens* herangezogen werden, unterscheiden sich auch die jeweils angelegten Antwortskalen in der Auswahl der Antwortkategorien bzw. der Skalenbreite. So setzt der WVS auf dichotome Antwortmöglichkeiten, während der GSS drei Antwortkategorien zu Auswahl stellt. Der ESS setzt hingegen auf eine 11stufige Bewertungsskala. Während REESKENS/HOOGHE und ZMERLI/NEWTON davon ausgehen, dass Ratingskalen, die möglichst viele Abstufungen enthalten, die beste Abbildung der Nuancierung von *generalisiertem Vertrauen* darstellen und eine größere Reliabilität aufweisen als dichotome Skalen, ist USLANER davon überzeugt, dass die Bewertung anhand zweier sich diametral gegenüberstehender Kategorien die bessere Variante darstellt. Für die Verwendung mehrstufiger Ratingskalen spricht, dass diese eine größere Varianz hinsichtlich des individuellen Vertrauenslevels der Befragten abbilden können. Außerdem ermöglicht das pseudo-metrische Skalenniveau von mehrstufigen Ratingskalen eine einfache Indexerstellung über mehrere Items per Summenscore. Zusätzlich spiegelt die Möglichkeit, sein Vertrauenslevel abzustufen zu können, die Realität *generalisierten Vertrauens* besser wider. Entsprechend einer dichotomen Antwortskala müsste Vertrauen entweder immer vorhanden oder immer abwesend sein. Tatsächlich ist das Vertrauen der Menschen gegenüber anderen oft viel nuancierter. Dichotome Antwortskalen würden daher einen Informationsverlust mit sich bringen. (Reeskens/Hooghe (2008), S. 520; Uslaner (2010), S. 3ff; Uslaner (2012b), S. 76ff; Zmerli/Newton (2008), S. 714)

4. Messung generalisierten Vertrauens

Eine hohe Anzahl an Abstufungen innerhalb einer Ratingskala kann dabei allerdings auch negative Auswirkungen auf das Antwortverhalten haben. USLANER führt unter Berufung auf MILLER und SCHWARTZ an, dass eine Skala mit mehr als sieben Antwortoptionen eher zur Konfusion bei den Befragten führt, als dass sie eine Verbesserung der Messpräzision bewirkt. Höherwertige Antwortskalen sorgen vielmehr dafür, dass die Befragten die Fähigkeit verlieren, zwischen den Antwortalternativen zu unterscheiden. USLANER verweist in diesem Zusammenhang auch auf das Phänomen des „*clumping*“ bzw. einer Tendenz zur Mittelkategorie. Beim Bearbeiten von Ratingskalen mit einer Vielzahl von Antwortoptionen tendieren die Befragten dazu, ihre Antworten um die Mittelkategorie der Skala herumanzuordnen. Eine viele Optionen umfassende Ratingskala trennt somit nicht gut zwischen vertrauensvollen und misstrauischen Personen. Nach USLANER handelt es sich bei *clumping* außerdem um einen Ausdruck sozial erwünschten Antwortverhaltens. *Clumping* geschieht hiernach systematisch, wobei insbesondere Personen mit einem niedrigen Vertrauenslevel dazu neigen, die sozial mehr akzeptierte neutrale Mittelkategorie auszuwählen als eine Option am unteren negativen Ende der angelegten Skala. USLANER führt indes mehrere Gründe für die Beibehaltung der Standarditems mit dichotomer Antwortskala an. Erstens verstehen die Befragten die antagonistische Polung des dichotomen Antwortformats besser. Zweitens haben Paneluntersuchungen gezeigt, dass die Messungen mittels der dichotomisierten Standarditems über die Zeit stabil bleiben. Drittens würde die Veränderung der Items bzw. die Anwendung einer anderen Messmethode die Möglichkeit der retrospektiven Vergleichbarkeit der Vertrauenslevel zerstören. Eine Kontinuität der Messung wäre nicht mehr gegeben. (Miller (1956); Schwartz (2003); Uslaner (2007), S. 9, S. 15; Uslaner (2010), S. 2f, S. 6ff, S. 21; Uslaner (2012b), S. 77f; Uslaner (2012c), S. 102f, S. 111)

Auch einer Erweiterung der bestehenden Standardskala zur Messung *generalisierten Vertrauens* um das Item „*Generally speaking, how often can you trust other people?*“, mit den Ausprägungen „*most of the time*“, „*about half of the time*“, „*once in a while*“ und „*never*“ aus der ANES Pilot Study (American National Election Study) lehnt USLANER ab, da auch hier das Problem des *clumping* auftritt sowie starke Diskrepanzen zu den Ergebnissen der Standarditems bestehen. Dabei können die Abweichungen der Vertrauensangaben in den Standarditems von denen in diesem Zusatzitem als weiterer Hinweis für den Einfluss sozialer Erwünschtheit auf die Erfassung *generalisierten Vertrauens* aufgefasst werden. So fand USLANER beispielsweise, dass Befragte, die angegeben hatten, Menschen immer zu vertrauen, 20% der Meinung waren, dass man nicht vorsichtig genug sein kann beim Umgang mit anderen. (Uslaner (2007), S. 7ff; Uslaner (2012c), S. 103ff)

5. Hypothesen

Im nachfolgenden Abschnitt sollen die Forschungsziele und zu prüfenden Hypothesen der vorliegenden empirischen Untersuchung vorgestellt werden. Das Forschungsvorhaben gliedert sich dabei in eine Pilotstudie und eine Hauptstudie. Im Rahmen der Pilotstudie werden zwei Vertrauensskalen entlang der Designvorgaben für sogenannte Objektive Tests entwickelt, während die Hauptstudie ein Surveyexperiment enthält, das diese Vertrauensskalen mit der bisherigen Standardmethode zur Erfassung Generalisierten Vertrauens kontrastiert. Dabei soll insbesondere geprüft werden, ob durch die Verwendung der im Folgenden als Objektiv-Skalen bezeichneten Messverfahren eine im Vergleich zur Standardmethode verbesserte Messgüte hinsichtlich der Zieldimension „Generalisiertes Vertrauen“ herbeigeführt werden kann.

5.1 Integration von direkten und indirekten Messverfahren

Ziel der Pilotstudie ist es, ein Messinstrument zu entwickeln, das die Eigenschaften direkter und indirekter Messverfahren in sich vereint. Als Ansatzpunkt für eine solche Hybridklasse wurde dabei die als Fragebogentyp bezeichnete Variante Objektiver Tests gewählt. Diese Kategorie Objektiver Tests bietet sich, vor allen anderen indirekten Messverfahren, aufgrund ihrer konzeptionellen Nähe zu direkten Messverfahren und im Speziellen zu sogenannten Selbstbeurteilungs- bzw. Selbstauskunftsmaßen als Basis für einen solchen Hybrid an. Die Konzeption einer Hybridklasse in Form eines Fragebogens ermöglicht außerdem eine Adaption auf eine Vielzahl von Gegenstandsbereichen, eine spezifische Nachfrage nach bestimmten Merkmalen sowie die Erhebung auch großer Stichproben. Daneben vereinfacht das Fragebogendesign die Bestimmung von Validität und Reliabilität des Messinstruments. Zudem stellen direkte Messverfahren, die die Form qualitativer oder quantitativer Interviews bzw. Befragungen annehmen, die mit Abstand meistgenutzten Messinstrumente innerhalb der empirischen Sozialforschung dar. Der Entwurf einer Hybridklasse als Fragebogen würde insofern eine Weiterentwicklung dieser nahezu universal einsetzbaren Methode empirischer Sozialforschung bedeuten. (Lucas/Baird (2013); Mummendey/Grau (2014), S. 19f; Petty/Cacioppo (1996), S. 22)

Die Ähnlichkeit des hybriden Messinstruments mit direkten Messverfahren und im Speziellen mit Selbstoffenbarungsmaßen ergibt sich im Wesentlichen dadurch, dass es sich aus einer Likert-Skala zusammensetzt, die Items enthält, die eine einzelne Dimension messen und anhand von Ratingskalen bewertet werden sollen. (Bortz/Döring (2016), S. 269) Übereinstimmungen mit indirekten Messverfahren und insbesondere mit Objektiven Tests ergeben sich durch die angestrebte Intransparenz der Messintention, die Vermeidung einer Selbstbeurteilung und die Datenerhebung innerhalb einer standardisierten Situation. (Schmidt (1975), S. 19)

Erreicht wird die Verschleierung der Messintention bzw. die mit der Zieldimension nicht übereinstimmende Augenscheinvalidität der Fragebogenitems durch eine Kombination verschiedener Maßnahmen. So wird dem Fragebogen zunächst eine Coverstory²⁷ vorangestellt,

²⁷ „Cover Story, eine falsche, aber plausible Erklärung für den Sinn und Zweck einer Untersuchung oder eines Experiments gegenüber den Vpn. Dahinter verbirgt sich die Absicht, den Einfluss von Hinweisen aus der exp. Situation zu kaschieren oder zumindest einzuschränken.“ (Wirtz/Strohmer (2013), S. 342)

5. Hypothesen

die von der eigentlichen Zieldimension ablenken bzw. über die Zielsetzung der Studie hinwegtäuschen soll. Dies geschieht in erster Linie, um die geringe Augenscheinvalidität der verwendeten Items nicht zu unterminieren und um Erwartungseffekte auszuschließen. (Bortz/Döring (2016), S. 125f, S. 197) Das Messinstrument ist in diesem Sinne in einer Art von Blindstudie eingebettet. Im Gegensatz zu klassischen Blind- oder Doppelblindstudien, bei denen entweder nur dem Versuchsleiter oder sowohl Versuchsleiter als auch Probanden die zu prüfende Hypothese nicht bekannt ist, werden hier allerdings ausschließlich die Versuchspersonen über das Forschungsziel im Unklaren gelassen. (Bortz/Döring (2016), S. 196ff; Diekmann (2011), S. 338)

Die geringe Augenscheinvalidität der Items hinsichtlich der Zieldimension wird dadurch bewirkt, dass die verwendeten Statements keinen direkten Bezug zur Zieldimension aufweisen, sondern versuchen das jeweilige Konstrukt anhand von Korrelaten sowie verwandten Aspekten und assoziiertem Verhalten indirekt zu erfassen. Die Items sind deshalb so formuliert, dass die verwendeten Statements weder das zu messende Konstrukt enthalten bzw. nennen, noch dass deren Wortlaut einen expliziten Hinweis auf die Zieldimension gibt. Das methodische Vorgehen entspricht somit einer indirekten Fragestrategie (Diekmann (2011), S. 474ff), wobei sich das Item-Design an den von CATTELL entwickelten Objektiven Tests des Fragebogentyps orientiert. So zeigen verschiedene Studien, dass insbesondere die Intransparenz der Messintention Objektiver Tests zur Reduktion bewusster Modifikationen bzw. Verfälschungen des individuellen Antwortverhaltens beitragen. (Arendasy/Sommer/Herle/Schützhofer/Inwanschitz (2011); Elliot/Lawty-Jones/Jackson (1996); Ziegler/Schmidt-Atzert/Bühner/Krumm (2007)) Beispiele für diese Variante Objektiver Tests sowie die jeweiligen Mess- und Auswertungsprinzipien lassen sich bei SCHMIDT und CATTELL/WARBURTON²⁸ finden. (Schmidt (1975); Cattell/Warburton (1967)) Nachfolgend sollen zwei dieser Beispiele, die im Rahmen der Erfassung von Vertrauen/Misstrauen zum Einsatz kommen, vorgestellt werden:

T8 Criticalness of Evaluation (Cattell/Warburton (1967), S. 305f)

Goodness of Work

Theory

The theory has to do with two distinct performances – speed of social type of judgement and criticalness of judgement. The former is a byproduct from several types of measurement. High exuberance (U.I. 21) and lack of inhibition (U.I. 17-) should be principally involved in high social judgement speed on serious topics.

Criticalness can run from hostility-aggression at one end of the scale, to exactingness and the possession of high standards at the other. Different tests are designed to bring out these extremes. In T8 criticalness is assessed by actually presenting examples of performance in

²⁸ Eine Übersicht über Objektive Tests, die von CATTELL zur Erfassung von Misstrauen bzw. Vertrauen („Timid Distrust (Inhibition-Timidity) vs. Trustingness“) entwickelt wurden, lässt sich bei CATTELL/WARBURTON finden. (Cattell/Warburton (1967), S. 197, S. 280f) Objektive Tests im Fragebogenformat, deren theoretischer Hintergrund, die jeweilige Implementation bzw. Handhabung und die zugehörige Auswertungsstrategie sind dabei auf den folgenden Seiten abgedruckt: S. 301f, S. 305f, S. 319f, S. 321f; S. 323f, S. 329f, S. 345f, S. 515, S. 521, S. 568f, S. 573

different fields and having them evaluated. A properly designed measure here should catch the carping and subtly hostile criticalness of the unresponsive, defensive person who is extreme on U.I. 29(-) and to some extent the overinhibited (U.I. 17) person.

Rationale

It is important that the examples (a) come from many fields, so that particular familiarity with the content of one will not bias the evaluation, and (b) have a middle answer as the truly correct one, since the tendency to give either overly severe or overly lenient responses will otherwise tend to be confound with sheer low accuracy.

Test Extract

On the next page you will read about a number of performances by people of various ages and training. Rate each one by putting a check mark in the A, B, C, or D column at the right of the page indicating how good you think the performance is for a person of that age and education.

1. What would you say of a drug store that takes an average of 20 minutes to make up a prescription?
 - A.) Very good
 - B.) Good
 - C.) Poor
 - D.) Very poor

2. How good is the following, as an attempt by college student to divide exactly, by eye, the distance between X and Y? X. . Y.
 - A.) Very good
 - B.) Good
 - C.) Poor
 - D.) Very poor

Scoring

For each item, score the most favorable response 4, the next response 3, the next 2, and the least favorable response 1. Add these scores and divide by the number of items done

T8 Kritische Bewertung (Schmidt (1975), S. 25)

1. Wie gut ist die Leistung eines 30jährigen kaufmännischen Angestellten, der ohne vorheriges Training 1000 Meter in 4½ Minuten läuft?
 - a.) Sehr gut
 - b.) Gut
 - c.) Mittelmäßig
 - d.) Schlecht
 - e.) Sehr schlecht

5. Hypothesen

2. Wenn eine Hausfrau von 40 Jahren 25 Seiten eines Taschenbuch-Romans in 34 Minuten liest, ist das
- a.) Sehr gut
 - b.) Gut
 - c.) Mittelmäßig
 - d.) Schlecht
 - e.) Sehr schlecht

T16 Pessimism about Events (Cattell/Warburton (1967), S. 319f)

Judgement of Human Nature

Theory

A high score in this test contributes to Factor U.I. 33. The factor is one of pessimism and caution. The content of the test includes suspiciousness, misanthropy, unwillingness to take a risk, and shrewd realism of a New England or Scottish kind.

Rationale

This test comprises a series of items concerning the effect of various social and moral actions. The plan is to present as many brief items of this kind as can be completed in 5 minutes. S answers in either a pessimistic or an optimistic direction.

Design

It was decided not to include a „don't know“ category of answer, since the pessimist might regard knowing nothing as being even more cautious than being pessimistic. Items are therefore forced choices, with „agree“ and „disagree“ categories only.

Test Extract

What to do: This is a test of judgement about people. Indicate by checking the space after „agree“ or „disagree“ what your judgement is on the following psychological statements.

1. Most serious crimes could be eliminated if the juvenile offenders could be given a new start by a happy home life and decent living conditions.
 - a.) Agree
 - b.) Disagree
2. For every act of kindness you do, you receive an equal amount of kindness from others.
 - a.) Agree
 - b.) Disagree

Scoring

Proportion of „pessimistic“ items checked according to key.

5. Hypothesen

T16 Pessimismus gegenüber Ereignissen (Schmidt (1975), S. 26)

1. Kinder können sich im Leben nur bewähren, wenn man sie streng erzieht.
 - a.) Stimme zu
 - b.) Stimme nicht zu
2. Wenn jemand seinen Geldbeutel auf dem Tisch eines Restaurants vergißt, wird er in 95 von 100 Fällen zurückgegeben.
 - a.) Stimme zu
 - b.) Stimme nicht zu

Eine Selbstbeurteilung der Probanden wird dadurch vermieden, dass die im Rahmen des Messinstruments eingesetzten Items ausschließlich generalisierte Andere betreffen. Die Probanden machen demnach keine Aussage über sich selbst, sondern evaluieren das Verhalten Dritter, bestimmen den Grad der Gültigkeit allgemeiner Aussagen über die menschliche Natur, entscheiden über die Allgemeingültigkeit von gegenstandsspezifischen Normen und Werten oder urteilen über situationsspezifisch idealtypisches Verhalten. Dabei sollten die verwendeten Statements immer einen indirekten Zusammenhang, in Einzelfällen auch direkten Zusammenhang, mit der jeweiligen Zieldimension aufweisen bzw. sich als konkretes Handeln oder verwandte Dimension aus dem Zielkonstrukt ableiten lassen. Typische Items können dabei die folgende Form annehmen und werden im Kontrast zu klassischen Fragebogenitems dargestellt:

Tab. 5.1: Beispiele: Klassische Items vs. Objektiv-Items (Zieldimension: Generalisiertes Vertrauen)

„Vertrauensbereitschaft“	
Klassisches Item	„Ganz allgemein gesprochen: Glauben Sie, dass man den meisten Menschen vertrauen kann, oder dass man im Umgang mit anderen Menschen nicht vorsichtig genug sein kann?“
Ausprägungen	<i>Man kann vertrauen. ... Man kann nicht vorsichtig genug sein.</i>
Objektiv-Item	„Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.“
Ausprägungen	Stimme nicht zu...Stimme zu
„Hilfsbereitschaft“	
Klassisches Item	„Und glauben Sie, dass die Menschen versuchen hilfsbereit zu sein, oder dass die Menschen meistens auf den eigenen Vorteil bedacht sind?“
Ausprägungen	<i>Versuchen hilfsbereit zu sein. Sind auf den eigenen Vorteil bedacht.</i>
Objektiv-Item	„Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.“
Ausprägungen	Stimme nicht zu...Stimme zu
„Fairness“	
Klassisches Item	„Glauben Sie, dass die meisten Menschen versuchen Sie auszunutzen, wenn Sie die Gelegenheit dazu haben, oder versuchen die meisten Menschen, sich fair zu verhalten?“
Ausprägungen	<i>Versuchen einen auszunutzen. Versuchen sich fair zu verhalten.</i>
Objektiv-Item	„Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.“
Ausprägungen	Stimme nicht zu...Stimme zu

Das Item-Design weist dabei Ähnlichkeit mit der Technik des *Indirect Questioning* und der *nominative technique* auf bzw. macht sich deren theoretische Annahmen zu nutze. Die Methode des *Indirect Questioning* versucht sozialerwünschtes Antwortverhalten zu reduzieren, indem die Probanden „answer structured questions from the perspective of another person or group“ (Fisher (1993), S. 303; Fisher/Tellis (1998), S. 563). Eine nach diesem Prinzip designte indirekte Frage „asks respondents to make predictions about how a similar other would think or act in a particular situation“ (Fisher/Tellis (1998), S. 563). Es wird dabei davon ausgegangen, dass die

5. Hypothesen

Probanden ihre unbewussten Einstellungen auf die Befragungssituation projizieren und dadurch mit ihrer Antwort in Wahrheit ihre eigenen unverfälschten Überzeugungen und Gefühle preisgeben. Die vermeintliche Tatsache, Aussagen über die soziale Umwelt zu machen, suggeriert den Probanden, dass ihre Antworten eher faktenbasiert sind als auf individuellen Meinungen zu fußen. *Indirect Questioning* ermöglicht es den Probanden auf diese Weise somit „to describe their own feelings behind a facade of impersonality“ (Simon/Simon (1974), S. 586). Unklar bleibt allerdings, in welchem Umfang die Prognosen bezüglich anderer Personen Informationen über das Selbst der Probanden widerspiegeln, und ob die Probanden in anderer Art und Weise antworten würden, wenn sie selbst Adressat der gestellten Fragen wären. (Fisher (1993), S. 304; Fisher/Tellis (1998), S. 563) Allerdings zeigen die Studien von FISHER und FISCHER/TELLIS, dass die Probanden im Rahmen von *Indirect Questioning* tatsächlich zu einer Projektion der eigenen Überzeugungen und Einschätzungen auf ihr Antwortverhalten neigen und indirekte Fragen im Vergleich zu direkten Fragen weniger stark durch Effekte sozialer Erwünschtheit beeinflusst sind. Die Anfälligkeit für sozialerwünschtes Antwortverhalten wird dabei durch die soziale Nähe der Item-Referenzgruppe moderiert. So erhöht sich die Wahrscheinlichkeit sozialerwünschten Antwortverhaltens, wenn, anstatt von Aussagen über andere Menschen im Allgemeinen, Aussagen über Personen aus dem engen persönlichen Umfeld (z.B. beste Freunde) gemacht werden sollen. Außerdem zeigte sich, dass durch die Technik des *Indirect Questioning* eine bessere Schätzung des wahren Wertes insbesondere sozial-sensibler Merkmale und Eigenschaften ermöglicht wird. (Fisher (1993), S. 307, S. 309, S. 312; Fisher/Tellis (1998), S. 565f)

Der Unterschied zwischen dem Item-Design des *Indirect Questioning* und dem Item-Design der hier dargestellten Hybridklasse aus direkten und indirekten Messverfahren liegt im Wesentlichen in der Ausprägung der gegenstandsspezifischen Augenscheinvalidität. Während im Rahmen des *Indirect Questioning* die Zieldimension der Items aufgedeckt bleibt und sich lediglich der Bezugsrahmen der Items vom Selbst der Probanden zum generalisierten Anderen verlagert, zeichnen sich die Items im Kontext der Hybridklasse durch eine geringe bzw. mit der Messintention nicht übereinstimmende Augenscheinvalidität aus und realisieren gleichzeitig eine Veränderung der Referenzgruppe.

Auch die *nominative technique* wurde als Methode zur Reduktion sozialerwünschten Antwortverhaltens entwickelt und geht auf die sogenannten *multiplicity techniques* von SIRKEN (Sirken (1970)) zurück. Die *nominative technique* beschreibt dabei eine Messmethode, in deren Rahmen die Probanden keine Fragen bezüglich ihres eigenen Verhaltens, sondern Fragen bezüglich des Verhaltens anderer Personen, die ihnen bekannt sind (z.B. Freunde, Bekannte oder Nachbarn), beantworten sollen. Die Probanden fungieren somit als Informanten hinsichtlich potentiell sozial-sensiblen Verhaltens sogenannter „nominierter“ Personen. Auf diese Weise wird eine Schattenstichprobe generiert, wobei sowohl die Probanden als auch die Personen, über die Informationen eingeholt werden, anonym bleiben. Die hinter diesem Vorgehen stehende Idee ist, dass die Probanden ehrlicher und wahrheitsgetreuer antworten werden, wenn sie über abweichendes Verhalten anonymer Personen berichten, anstatt wenn sie über eigenes abweichendes Verhalten berichten. (Chaudhuri, A. (2011), S. 116ff; Chaudhuri/Christofides (2013), S. 22, S. 24f, S. 129ff; Krumpal (2013), S. 2040; Lee (1993), S. 86f; Scholl (2015), S. 224; St. John/Edwards-Jones/Gibbons/Jones (2010), S. 1026; Sudman/Blair/Bradburn/Stocking (1977), S. 261)

Ein Nachteil dieses Verfahrens ist, dass die *nominative technique* ausschließlich auf Aktivitäten angewendet werden kann, die potentiell beobachtbar sind. Private oder intime Handlungen und

5. Hypothesen

Verhaltensweisen sozialisolierter Personen können somit nicht erfasst werden. Außerdem gibt es keine Garantie dafür, dass die Probanden ihre Informanten-Rolle akzeptieren oder zu over- bzw. underreporting neigen. (Lee (1993), S. 87; Scholl (2015), S. 222, S. 224; Sudman/Blair/Bradburn/Stocking (1977), S. 262) Ihre Wirksamkeit hat die *nominative technique* vornehmlich in Studien zur Verbreitung von Drogenmissbrauch bewiesen, indem sie ein under-reporting verhinderte. (Lee (1993), S. 88; Miller (1985); Sudman/Blair/Bradburn/Stocking (1977))

Im Unterschied zur Fragetechnik der Hybridklasse und zum *Indirect Questioning* nimmt die *nominative technique* die Aussagen der Probanden über andere Personen allerdings nicht als Surrogat für das Verhalten der Probanden an. Vielmehr soll eine Erhöhung der Stichprobengröße (Schattenstichprobe) bewirkt werden, um auf diese Weise ein vollständigeres Gesamtbild über die Verteilung einer spezifischen Verhaltensform innerhalb der Population zu erhalten. (Lee (1993), S. 87; Sudman/Blair/Bradburn/Stocking (1977), S. 261) Gemeinsam ist allen drei Fragetechniken dagegen, dass versucht wird, anhand eines indirekten Zugangs Informationen über Merkmale, Eigenschaften und Verhalten von Personen zu erhalten, um auf diese Weise reaktive Effekte zu minimieren.

Die Erfassung in einer standardisierten Situation wurde im Rahmen der Hybridklasse schließlich durch die Verwendung eines standardisierten Fragebogens als Messinstrument umgesetzt. Auf diese Weise wird jeder Proband mit der gleichen Abfolge von Items konfrontiert und der inhaltliche Stimulus konstant gehalten, was in einer Äquivalenz der Interviewsituation für alle Probanden resultiert und eine Vergleichbarkeit der Messergebnisse ermöglicht. (Schnell/Hill/Esner (2008), S. 323)

Neben den hier dargestellten Gemeinsamkeiten mit Objektiven Tests bzw. Adaptionen der Vorgaben für Objektive Tests ergeben sich auch eine Reihe von Unterschieden zwischen Objektiven Tests und Hybridklasse. Die wichtigste Abweichung stellt dabei die Tatsache dar, dass im Rahmen der Hybridklasse keine Leistungsaufgaben zum Einsatz kommen und die Ausprägungen der Merkmale und Eigenschaften der Probanden nicht anhand deren Performanz abgeleitet werden. Auch erfolgt die Messung nicht anhand von standardisierten Miniatursituationen, die realen Lebensumständen nachempfunden sind, noch wird explizit das Verhalten der Probanden erfasst. Die Verwendung von Ratingskalen zur Bewertung der Fragebogenitems hat außerdem zur Folge, dass der Grundsatz der Vermeidung einer Selbstbeurteilung der Probanden zumindest teilweise verletzt wird und auch das zugrundeliegende Messprinzip nicht vollständig intransparent bleibt. So vollziehen die Probanden bei der Auswahl einer Antwortalternative eine bewusste Selektion und auch die Wertigkeit der Antwortoption ist für die Probanden erkennbar. Unklar bleibt den Probanden allerdings, welche Zieldimension die jeweiligen Items abbilden, welcher Ausprägung auf der Zieldimension ihre Antwort entspricht und wie ihre Antwort im Rahmen der Datenauswertung verarbeitet wird. Auch der Einsatz einer Coverstory ist im Rahmen Objektiver Tests nicht vorgesehen. Vielmehr darf mit Hilfe der Testinstruktionen keine Irreführung der Probanden herbeigeführt werden.

(Ortner/Horn/Kersting/Krumm/Kubinger/Proyer/Schmidt-Atzert/Schuhfried/Schütz/Wagner-Menghin/Westhoff (2007), S. 62)

Es ergibt sich an dieser Stelle zunächst folgende Fragestellung, die den explorativen Charakter des Forschungsziels dieses Forschungsprojektes unterstreicht:

„Lässt sich eine ausreichend reliable Skala erstellen, die eine valide Erfassung des theoretischen Konstruktes „Generalisiertes Vertrauen“ ermöglicht und den hier aufgestellten Kriterien einer Hybridklasse aus direkten und indirekten Messverfahren²⁹ (mit der Zieldimension gering bzw. nicht übereinstimmende Augenscheinvalidität der Items und Vermeidung einer Selbstbeurteilung durch Bezug auf generalisierte Andere) entspricht?“

5.2 Homogenität und Reliabilität

Im Allgemeinen unterscheidet man homogene und heterogene Tests. Ein homogener Test beschreibt in der Regel eine eindimensionale Skala bzw. eine Skala, die nur eine einzige Dimension abbildet. Die in einer solchen Skala zusammengefassten Items erfassen dabei dasselbe Merkmal oder repräsentieren ähnliche Merkmalsfacetten. Ein heterogener Test hat dagegen einen breiteren Geltungsbereich und setzt sich aus Items zusammen, die unterschiedliche Merkmalsfacetten oder verschiedene Merkmale erfassen. (Bühner (2011), S. 178; Fisseni (1997), S. 54f; Lienert/Raatz (1998), S. 36; Moosbrugger/Kelava (2012), S. 33)

„Homogenität bedeutet dabei inhaltliche Einheitlichkeit bei vollkommen erhaltener formaler Unabhängigkeit der einzelnen Aufgaben voneinander. (...) Solche homogenen Tests erfassen gewöhnlich nur ein eng umschriebenes Persönlichkeitsmerkmal, (...); sie haben meist eine relativ hohe Reliabilität, und man kann daher mit verhältnismäßig wenigen Aufgaben auskommen.“ (Lienert/Raatz (1998), S. 36)

„Unter Heterogenität von Aufgaben versteht man ihre inhaltliche Vielgestaltigkeit, wie sie zweckmäßigerweise bei Tests angestrebt wird, die ein schwer abgrenzbares Persönlichkeitsmerkmal oder einen ganzen Merkmalskomplex erfassen sollen.“ (Lienert/Raatz (1998), S. 36)

Die Items der Objektiv-Skalen sollen nach Möglichkeit nur die einzelne Dimension „Generalisiertes Vertrauen“ abbilden. Es ergibt sich daher folgende Hypothese:

H1: „Die Objektiv-Skala entspricht einem homogenen Test und zeichnen sich durch Eindimensionalität aus bzw. bildet nur eine einzige Dimension ab.“

Homogene Tests aufgrund ihrer Konsistenz bzw. der Zusammenfassung inhaltsähnlicher Items normalerweise oft hoch reliabel. Dabei stellt Reliabilität neben Objektivität und Validität eines der zentralen Gütekriterien quantitativer Messungen dar. (Diekmann (2011), S. 247ff) Mit Reliabilität bzw. Zuverlässigkeit wird die Messgenauigkeit eines Tests beschrieben, wobei irrelevant ist, ob der fragliche Test auch tatsächlich eine valide Messung der angepeilten Zieldimension vollzieht. Im Fokus der Reliabilitätsanalyse steht somit nicht die Frage danach, was ein Test misst, sondern wie präzise er das, was er misst, abbildet. So ist es möglich, dass ein Test zwar eine exakte Messung erlaubt, dabei aber u.U. ein vollkommen anderes als das erwünschte Konstrukt erfasst. (Bühner (2006), S. 35f; Bühner (2011), S. 60f; Fisseni (1997), S. 70; Lienert/Raatz (1998), S. 9f; Moosbrugger/Kelava (2012), S. 120; Mummendey/Grau (2014), S. 100ff)

²⁹ Anm. d. Verf.: Die Erhebung in einer standardisierten Situation ist bereits durch die Wahl des Fragebogen-Designs gewährleistet.

5. Hypothesen

„Unter Reliabilität wird die Genauigkeit einer Messung verstanden. Ein Testverfahren ist perfekt reliabel, wenn die damit erhaltenen Testwerte frei von zufälligen Messfehlern sind. Das Testverfahren ist umso weniger reliabel, je größer die Einflüsse von zufälligen Messfehlern sind.“ (Moosbrugger/Kelava (2012), S. 120)

Voraussetzung für die Bestimmung der Reliabilität eines Tests sind standardisierte Messbedingungen bzw. die Objektivität des Messinstruments, der Testsituation, der Testdurchführung und der Testauswertung. Die Objektivität der Objektiv-Skalen ergibt sich aus deren Design als standardisierter Fragebogen. Allen Probanden wird auf diese Weise die gleiche Messvorschrift vorgelegt, sodass eine Vergleichbarkeit der Antworten der verschiedenen Probanden garantiert ist.

Hinsichtlich der Messgenauigkeit der Objektiv-Skalen ergibt sich nun folgende Hypothese:

H2: „Die Objektiv-Skala ermöglicht eine präzise Messung Generalisierten Vertrauens bzw. bei der Objektiv-Skala handelt es sich um ein ausreichend reliables Messinstrument.“

5.3 Validität

„Unter Validität (Gültigkeit) eines Messinstrumentes versteht man das Ausmaß, in dem das Messinstrument tatsächlich das misst, was es messen sollte.“ (Schnell/Hill/Esser (2008), S. 154)

Validität stellt neben Objektivität und Reliabilität eines der Hauptgütekriterien quantitativer empirischer Messungen dar und „befasst sich mit der inhaltlichen Übereinstimmung zwischen dem vom Test gemessenen Merkmal und dem Merkmal, das man messen will“ (Moosbrugger/Kelava (2012), S. 13). Dabei gilt ein Messinstrument, das nicht dazu in der Lage ist, das intendierte Zielkonstrukt abzubilden, oder das eine von der Messintention abweichende Zieldimension erfasst, als inadäquat und unbrauchbar, eine gültige Messung des jeweiligen interessierenden Merkmals zu gewährleisten.

Hinsichtlich der Validität der Objektiv-Skala ergibt sich daher folgende grundlegende Hypothese:

H3: Bei der Erfassung von Generalisiertem Vertrauen mittels der Objektiv-Skala handelt es sich um eine valide Messung bzw. die Objektiv-Skala und die Standard-Skala sind inhaltlich äquivalent.

Im Rahmen einer Validierung lassen sich im Allgemeinen die folgenden drei Subkategorien unterscheiden: Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität. (Bortz/Döring (2016), S. 445ff; Bühner (2011), S. 61ff; Moosbrugger/Kelava (2012), S. 144ff; Schnell/Hill/Esser (2011), S. 155ff)

„Der Begriff der *Inhaltsvalidität* bezieht sich darauf, inwieweit die Inhalte eines Tests bzw. der Items, aus denen er sich zusammensetzt, tatsächlich das interessierende Merkmal erfassen.“ (Moosbrugger/Kelava (2012), S. 148)

Als inhaltsvalide gelten Items demnach, wenn sie repräsentativ für das zu erfassende Merkmal sind, nach Möglichkeit sämtliche Aspekte des interessierenden Merkmals berücksichtigen und

5. Hypothesen

das Merkmal hinreichend präzise erfassen. (Bühner (2011), S. 61f; Mummendey/Grau (2014), S. 102f; Schnell/Hill/Esler (2008), S. 155)

Grundlegend für die Bestimmung der Inhaltsvalidität ist die Annahme eines homogenen Item- bzw. Indikatorenuniversums in dem alle Items enthalten sind, die zur Erfassung des fraglichen Merkmals verwendet werden können. Die zur Erfassung des Merkmals eingesetzten Items stellen dann eine repräsentative Stichprobe aus diesem Itemuniversum dar. Ob die ausgewählten Items tatsächlich repräsentativ für das zu messende Merkmal sind, wird in der Regel auf Basis logischer und theoretischer Überlegungen bestimmt (*logische Validität*). (Bühner (2011), S. 62; Moosbrugger/Kelava (2012), S. 15; Schnell/Hill/Esler (2008), S. 133f) Die Inhaltsvalidität ist insofern bereits in der theoretischen Beschreibung des Zielkonstruktes enthalten bzw. lässt sich aus dessen theoriegeleiteter Operationalisierung schlussfolgern. Schnell/Hill/Esler (2008), S. 155)

Hinsichtlich der Inhaltsvalidität der Objektiv-Skala lässt sich dann die nachfolgende Hypothese formulieren:

H3a: Die zur Objektiv-Skala zusammengefassten Items ermöglichen eine Abbildung der Dimension „Generalisiertes Vertrauen“ bzw. lassen eine präzise Messung des angestrebten theoretischen Konstruktes zu. Die Objektiv-Skala weist Inhaltsvalidität auf. [Inhaltsvalidität]

Objektive Tests sind dadurch definiert, dass sie das Verhalten bzw. die Einstellungen eines Individuums (1) in einer standardisierten Situation erfassen, (2) ohne dass der/die Befragte eine Selbstbeurteilung vornimmt und (3) ohne, dass das Messinstrument Hinweise auf die Messintention gibt. (Fisseni (1997), S. 309; Fisseni (2004), S. 210f; Schmidt (1975), S. 19) Wie bereits beschrieben, wurden diese hier genannten Charakteristika auch für die aus dem Fragebogentyp der Objektiven Tests abgeleitete Hybridklasse aus direkten und indirekten Messverfahren adaptiert.

Die Erfassung in einer standardisierten Situation stellt dabei eine Frage der Objektivität der Messung dar und wurde durch die Wahl des Untersuchungsinstruments umgesetzt. So zeichnet sich ein standardisierter Fragebogen dadurch aus, dass die Testsituation, die Testdurchführung und die Testauswertung normiert sind. Daneben ist Objektivität eine notwendige Voraussetzung für eine ausreichend hohe Skalenreliabilität. (Moosbrugger/Kelava (2012), S. 120)

Eine Selbstbeurteilung der Probanden wurde dadurch vermieden, dass die verwendeten Items ausschließlich auf generalisierte Andere bezogen waren und somit keine Urteile oder Aussagen bezüglich der Person der Probanden erforderlich machten.

Die Verschleierung der Messintention spiegelt sich in der sogenannten Augenscheinvalidität (*Face-Validity*) wider, die einen Teilaspekt der Inhaltsvalidität darstellt bzw. an diese angelagert ist. In der Regel sollte die Augenscheinvalidität eines Tests möglichst hoch ausfallen, da die Befragten andernfalls die Messintention des Tests nicht erkennen und sich keine stabilen Ergebnisse bzw. ein konsistentes Antwortverhalten einstellt. Den Befragten wird somit aufgrund der Itemformulierung und den Testinstruktionen unmittelbar klar, welche Zieldimension mit dem vorliegenden Untersuchungsinstrument erfasst werden soll bzw. dass ein Zusammenhang zwischen Testaufgaben und gemessenem Verhalten besteht. Dabei wird inhaltsvaliden Tests auch oft Augenscheinvalidität zugesprochen. (Bortz/Döring (2016), S. 445; Bühner (2011), S. 62) Entsprechend ergibt sich für die Augenscheinvalidität folgende Definition:

5. Hypothesen

„Augenscheinvalidität gibt an, inwieweit der Validitätsanspruch eines Tests, vom bloßen Augenschein her einem Laien gerechtfertigt erscheint.“ (Moosbrugger/Kelava (2012), S. 15)

Vorgabe für die hier entwickelte Hybridklasse ist dagegen, dass die Augenscheinvalidität entweder möglichst niedrig oder vollkommen eliminiert sein sollte. Hintergrund ist die Vermeidung subjektiver Verfälschungen durch die Befragten, die sich aus der Kenntnis der Messintention ergeben können. Ein Objektiver Test ist dann in diesem Sinne „objektiv“, wenn die Befragten aus den Testaufgaben nicht erkennen können, welches Konstrukt der Test abbilden soll, und demnach das Testresultat nicht bewusst manipulieren können. (Moosbrugger/Kelava (2012), S. 31)

Problematisch bleibt bei einer geringen Augenscheinvalidität allerdings, dass man nicht sicher sein kann, ob die Antworten der Befragten der intendierten aber verschleierte Zieldimension des Tests entsprechen oder ob sich Antworten der Befragten lediglich auf ein verwandtes Konstrukt der tatsächlichen Zieldimension oder ein vollkommen anderes Konstrukt beziehen. (Bungard (1984), S. 116ff; Hundleby (1973), S. 68; Orne (1969), S. 147ff; Schnell/Hill/Esler (2008), S. 220) Diesem Aspekt wird begegnet, indem die Entwicklung des Untersuchungsinstruments auf einer breiten sowohl empirisch als auch theoretisch fundierten Basis erfolgt. So kommen neben Items, die aus bereits validierten Fragebögen zur Erfassung von Vertrauen entnommen wurden und den Anforderungen der Hybridklasse entsprechen, auch Neuformulierungen, die sich aus den verschiedenen theoretischen Ansätzen zur Erklärung der Genese „Generalisierten Vertrauens“ ergeben, zum Einsatz.

Dies führt zu folgender Hypothese bezüglich der Augenscheinvalidität (*Face-Validity*) der Objektiv-Skala:

H3b: Es lässt sich für die Probanden per Augenschein kein unmittelbarer Zusammenhang zwischen den Items der Objektiv-Skala und der Zieldimension „Generalisiertes Vertrauen“ ableiten bzw. die Messintention der Objektiv-Items ist für die Probanden nicht zu durchschauen. [Augenscheinvalidität]

Mit anderen Worten: Die Objektiv-Items weisen hinsichtlich des Konstruktes „Generalisiertes Vertrauen“ eine geringe Augenscheinvalidität auf bzw. es besteht keine mit der Messintention übereinstimmende Face-Validity.

Die Konstruktvalidität stellt das zentrale Validitätskriterium innerhalb der Testtheorie dar und geht davon aus, dass es sich bei den mit einem Test erfassten Merkmalen immer um theoriegestützte latente Konstrukte handelt. Aus der theoretischen Verfasstheit der Konstrukte lassen sich dann empirisch überprüfbare Aussagen über Verbindungen zu anderen Konstrukten ableiten. Hierzu werden zunächst Indikatoren bestimmt, mit denen sich das nicht direkt beobachtbare Konstrukt laut den theoretischen Vorgaben abbilden lässt. Im Anschluss wird geprüft, ob die in einer empirischen Untersuchung gewonnenen Ergebnisse mit der Theorie übereinstimmen. Durch die Untersuchung der Zusammenhänge des fraglichen Konstruktes mit anderen Konstrukten wird dabei eine Vielzahl möglicher Operationalisierungen geprüft und es kann bestimmt werden, wie die Ergebnisse des neuen Tests bewertet und interpretiert werden müssen. (Bühner (2011), S. 63f; Moosbrugger/Kelava (2012), S. 145f, S. 153ff; Mummendey/Grau (2014), S. 106ff; Schnell/Hill/Esler (2008), S. 155ff) Im Wesentlichen soll aus der Konstruktvalidität abgeleitet werden, ob „der Test auch die Eigenschaft oder Fähigkeit misst, die er messen soll“ (Bühner (2011), S. 63).

5. Hypothesen

Hierbei werden drei verschiedenen Formen der Konstruktvalidität unterschieden. Die *konvergente Validität* wird anhand von Korrelationen mit Tests bestimmt, die das gleiche Merkmal oder ähnliche Merkmale erfassen. So wird im Sinne des Konzepts des Indikatorenuniversums und des Konzepts der multiplen Indikatoren davon ausgegangen, dass verschiedene Indikatoren desselben Konstruktes wechselseitig austauschbar sind, und es werden dementsprechend hohe Zusammenhänge zwischen beiden Messungen erwartet. (Bühner (2011), S. 64; Moosbrugger/Kelava (2012), S. 156ff; Schnell/Hill/Esner (2008), S. 158)

Die *diskriminante* oder *divergente Validität* wird anhand von Korrelationen mit Tests ermittelt, die andere Merkmale als die Zieldimension erfassen. Gezeigt werden soll, dass es sich bei dem zu prüfenden Test um eine diskrete Messung der Zieldimension handelt, die von der Messung des verwandten Konstruktes zu unterscheiden ist. Es werden dementsprechend niedrige Zusammenhänge erwartet. (Bühner (2011), S. 64; Moosbrugger/Kelava (2012), S. 156ff; Schnell/Hill/Esner (2008), S. 157f)

Im Rahmen der Bestimmung der *faktoriellen Validität* wird geprüft, ob die Zusammenhangsstruktur der verwendeten Items auf ein einzelnes latentes Konstrukt zurückgeführt werden kann. *Faktorielle Validität* ist somit dann gegeben, wenn die zur Messung der Zieldimension verwendeten Items ein gemeinsames eindimensionales Konstrukt abbilden und dieses von fremden Konstrukten unterschieden werden kann. *Faktorielle Validität* stellt dabei allerdings noch keinen hinreichenden Beweis für Konstruktvalidität dar, da auf diese Weise der Inhalt der Dimension noch nicht eindeutig bestimmt worden ist. (Bühner (2011), S. 64; Moosbrugger/Kelava (2012), S. 162; Schnell/Hill/Esner (2008), S. 161ff)

Auf die Bestimmung *divergenter* und *faktorieller Validität* wurde im Kontext der Hauptstudie verzichtet, da im eingesetzten Fragebogen keine Tests anderer Gültigkeitsbereiche enthalten waren bzw. die von Generalisiertem Vertrauen zu differenzierenden Variablen als externe Kriterien zur Bestimmung der Kriteriumsvalidität herangezogen wurden und die einzelnen Versuchsgruppen des Surveyexperimentes nicht genug Fälle enthielten, um eine Faktorenanalyse durchzuführen. Ein rudimentäre Bestimmung der *faktoriellen Validität* erfolgt allerdings im Rahmen der Pilotstudie.

Um die *konvergente Validität* zu ermitteln, wurde eine parallele Vertrauensmessung mit Hilfe der *Trust in People-Scale* durchgeführt. Die *Trust in People-Scale* entspricht dabei dem Standardverfahren zur Erfassung Generalisierten Vertrauens. (Kunz (2004), S. 204f; Robinson/Shaver/Wrightsmann (1991), S. 406ff)

Hinsichtlich der *konvergenten Validität* der Objektiv-Skala ergibt sich daher folgende Hypothese:

H3c: Die Items der Objektiv-Skala korrelieren hoch mit der Trust in People-Scale bzw. den Standarditems. Die Objektiv-Skala weist Konstruktvalidität auf. [Konstruktvalidität: Konvergente Validität]

Die *Kriteriumsvalidität* beschreibt den Zusammenhang zwischen der mit Hilfe des Messinstruments gemessenen Zieldimension und einem externen Kriterium. Sie dient in erster Linie dazu, Prognosen und Diagnosen hinsichtlich des Verhaltens und der Einstellungen der Befragten auf Basis der Testergebnisse abzusichern.

„Kriteriumsvalidität bedeutet, dass von einem Testergebnis auf ein für diagnostische Entscheidungen praktisch relevantes Kriterium außerhalb der Testsituation geschlossen werden kann. Kriteriumsvalidität

5. Hypothesen

kann durch empirische Zusammenhänge zwischen dem Testwert und möglichen Außenkriterien belegt werden. Je enger diese Zusammenhänge, desto besser kann die Kriteriumsvalidität als belegt gelten.“ (Moosbrugger/Kelava (2012), S. 164)

Es werden dabei verschiedene Formen der kriterienbezogenen Validitätsbestimmung unterschieden. Besondere Relevanz besitzen die *Vorhersagevalidität* (*prädiktive Validität*) und die *Übereinstimmungsvalidität* (*konkurrente Validität*). Erstere bezieht sich auf Korrelationen der Zieldimension mit Kriterien, die zeitlich später erhoben wurden. Hiermit soll bestimmt werden, ob sich auf Basis der Testwerte Prognosen hinsichtlich des Verhaltens der Probanden abgeben lassen. Der Zeitrahmen der Vorhersage ist dabei variabel. Letztere bezieht sich auf Korrelationen mit Kriterien, die zeitgleich mit der Zieldimension erfasst wurden. Hier stehen eher diagnostische Ziele im Vordergrund. So wird vom Testverhalten auf Verhalten außerhalb des Tests geschlossen. (Bortz/Döring (2016), S. 447; Bühner (2011), S. 63; Lienert/Raatz (1998), S. 223f; Moosbrugger/Kelava (2012), S. 165; Schnell/Hill/Esner (2008), S. 155f)

Als eher sekundär zu bewerten, sind die *retrospektive Validität* und die *inkrementelle Validität*. Erstere bezieht sich auf Zusammenhänge von Zieldimension und im Vorhinein ermittelten Kriterien, während sich letztere auf den Beitrag bezieht, den das neue Testinstrument leistet, um die Vorhersage eines Kriteriums über bestehende Testverfahren hinaus zu verbessern. (Bühner (2011), S. 63; Moosbrugger/Kelava (2012), S. 166)

Im Rahmen der Hauptstudie wird ausschließlich die *konkurrente Validität* bestimmt werden. Als externe Kriterien werden dabei das subjektive Wohlbefinden, das Vertrauen in Implementationsinstitutionen, die Eingebundenheit in informelle Netzwerke und die Mitgliedschaft in formelle Netzwerke herangezogen. Alle genannten Kriterien sollten dabei positiv mit Generalisiertem Vertrauen verbunden sein. Aus subjektivem Wohlbefinden resultiert in der Regel eine optimistische Weltwahrnehmung und die Mitmenschen werden als weniger bedrohlich empfunden. Als Konsequenz dieser Perzeption erhöht sich gleichzeitig die Bereitschaft in Vertrauensbeziehungen einzutreten. Die vergrößerte Vertrauensbereitschaft bezieht sich dabei gleichermaßen auf bekannte wie fremde Personen, sodass ein Anstieg des subjektiven Wohlbefindens auch eine Steigerung des Generalisierten Vertrauens nach sich zieht.

Durch Implementationsinstitutionen wird die Gültigkeit einer Rechtsordnung und institutioneller Regeln und damit der Bestand einer sozialen Ordnung gewährleistet. Das Handeln der verschiedenen Mitglieder einer Gesellschaft wird durch diese Normen gerahmt und kann dadurch vom Einzelnen besser vorausgesagt werden, was den Eintritt in Vertrauensbeziehungen weniger gefährlich macht.

Die Mitgliedschaft in informellen und formellen Netzwerken ermöglicht Kooperationserfahrungen. Je nachdem, ob diese positiv oder negativ besetzt sind, können diese sich förderlich oder hemmend auf die Entwicklung Generalisierten Vertrauens auswirken. So übertragen sich die Erfahrungen aus Vertrauensbeziehungen mit fremden Menschen in formellen Netzwerken oft auf den gesamtgesellschaftlichen Kontext, was in einem Anstieg Generalisierten Vertrauens resultieren kann.

Damit leitet sich in Bezug auf die Kriteriumsvalidität der Objektiv-Skala die folgende Hypothese ab:

H3d: Die Items der Objektiv-Skala korrelieren hoch mit relevanten externen Kriterien (Subjektives Wohlbefinden [Pilot- und Hauptstudie], Vertrauen in Implementationsinstitutionen, Eingebundenheit in informelle Netzwerke, Mitgliedschaft

in formellen Netzwerken [Hauptstudie]). Die Objektiv-Skala weist Kriteriumsvalidität auf. [Kriteriumsvalidität: Übereinstimmungsvalidität (konkurrente Validität)]

5.4 Differenz von hybriden Messverfahren und direkten Messverfahren

Explizites Ziel der Hauptstudie ist es, zu prüfen, ob sich zwischen den standardmäßig zur Erfassung Generalisierten Vertrauens verwendeten Items (*Trust in People-Scale* bzw. Standard-Skala) und den Objektiv-Skalen Messwertunterschiede ergeben und, ob diese potentielle Differenz auf den Einfluss sozialerwünschten Antwortverhaltens zurückzuführen ist. Insofern erweitert sich die zuvor formulierte Fragestellung wie folgt:

„Inwiefern wird durch die Verwendung der Objektiv-Skalen³⁰ eine Verbesserung der Messgüte gegenüber dem traditionellen Fragebogendesign bewirkt?“

Im Speziellen bedeutet dies:

*„Unterscheiden sich die Messwerte der aus der Pilotstudie resultierenden Objektiv-Skalen von den Messwerten der Standard-Skala (*Trust in People-Scale*) und ist dieser Effekt auf die Reduktion des Einflusses sozialerwünschten Antwortverhaltens infolge des speziellen Designs der Objektiv-Skalen zurückzuführen?“*

Zum Schluss soll außerdem überprüft werden, ob die Objektiv-Skalen einen Beitrag zur Lösung des Radiusproblems der Vertrauensmessung leisten können.

5.4.1 Reaktive Effekte

Wie bereits beschrieben, sind Objektive Tests im Wesentlichen dadurch gekennzeichnet, dass sie die interessierenden Aspekte eines Forschungsgegenstandes in (1) einer standardisierten Situation erfassen, (2) ohne dass die Probanden eine Selbstbeurteilung vollziehen und (3) ohne dass das Messinstrument über eine mit der Messintention übereinstimmende Augenscheinvalidität verfügt bzw. das Messinstrument Hinweise auf die jeweilige Zieldimension der Messung gibt. (Schmidt (1975), S. 19) Es sind diese Merkmale, die auch Anwendung im Design der Hybridklasse als Vereinigung von direkten und indirekten Messverfahren erfahren und einen Beitrag zur Reduktion reaktiver Effekte bzw. im Speziellen der Verringerung der Anfälligkeit für sozialerwünschtes Antwortverhalten und den daraus resultierenden potentiellen Messverfälschungen leisten sollen.

Um zu prüfen, ob die Designeigenschaften der Hybridklasse tatsächlich bewirken können, dass Messungen mit Hilfe der Hybridklasse im Kontrast zu Messungen mit klassischen direkten Messmethoden weniger stark durch Reaktivität verzerrt sind, werden die im Rahmen der Pilotstudie entwickelten Objektiv-Skalen dem Standardinstrument für die Messung von

³⁰ Anm. d. Verf.: Aus der Pilotstudie resultierten insgesamt zwei unterschiedlich lange Objektiv-Skalen. Die sogenannte Objektiv-Skala enthielt 14 Items, während sich die sogenannte Objektiv-Kurzskala aus 7 Items zusammensetzte, die auch schon Bestandteil der Objektiv-Skala waren.

5. Hypothesen

Generalisiertem Vertrauen (*Trust in People-Scale*³¹) gegenübergestellt. Dabei steht insbesondere die potentielle Beschränkung des Einflusses sozialerwünschten Antwortverhaltens durch die Anwendung der Objektiv-Skalen im Fokus des Forschungsinteresses.

Im Allgemeinen wird durch soziale Erwünschtheit das folgende Verhalten beschrieben:

„Sozial erwünscht‘ (social desirable) zu reagieren heißt entsprechend der Erwartung zu reagieren, daß die Art der Reaktion die Zustimmung einer sozialen Gruppe findet.“ (Mummendey (1981), S. 199)

Im Speziellen entspricht die absichtliche Verfälschung im Sinne sozialerwünschten Antwortverhaltens in der Regel der Simulation („*fake good*“) eines bestimmten Verhaltens. Das bedeutet: Es wird ein Verhalten vorgetäuscht, welches man sonst nicht zeigt. (Bortz/Döring (2016), S. 437ff; Bühner (2011), S. 125;) Die Probanden geben im Rahmen einer Messung somit nicht die wahre Ausprägung eines Merkmals oder ihren wahren Standpunkt gegenüber einem Sachverhalt bzw. Forschungsobjekt wieder, sondern weisen „a tendency to respond in a way that makes the respondent look good“ (Pauls/Stemmler (2003), S. 263) auf. Anders ausgedrückt:

„Motiviert durch die Furcht vor sozialer Verurteilung neigt man zu konformem Verhalten und orientiert sich in seinen Verhaltensäußerungen an verbreiteten gesellschaftlichen Normen und Erwartungen.“ (Bortz/Döring (2016), S. 437)

Sozialerwünschtes Antwortverhalten im Sinne von „*fake good*“ geht dabei in der Regel mit dem Versuch einher, besonders hohe Testwerte zu erreichen. (Bortz/Döring (2016), S. 436; Bühner (2011), S. 125) Hierbei ist allerdings zu beachten, dass je nach betrachtetem Merkmal nicht unbedingt allgemeinverbindliche Normen existieren müssen, sondern die Bestimmung eines gesellschaftlich erstrebenswerten Verhaltens bzw. sozialerwünschter Einstellungen von situationsspezifischen Variablen sowie der jeweiligen Bezugsgruppe abhängig sein können. Auch muss die Norm, an der sich das Antwortverhalten der jeweiligen Person orientiert, nicht unbedingt in direkter Abhängigkeit zum Inhalt des jeweiligen Items stehen.³² So kann mitunter auch die Bestrebung, möglichst niedrige Testwerte zu erzielen, Ausdruck sozialerwünschten Antwortverhaltens sein. Ein solches Antwortverhalten wird dann als Dissimulation („*fake bad*“) bzw. das Verschleiern oder Verbergen von Verhalten, das man normalerweise zeigt, bezeichnet. (Bühner (2011), S. 125; Mummendey/Grau (2014), S. 167f) Ob sich sozialerwünschtes Antwortverhalten in Form einer Simulation oder Dissimulation niederschlägt, ist nicht zuletzt auch davon abhängig, ob die Itempolung der sozialerwünschten Richtung entspricht. (Mummendey/Grau (2014), S. 168)

Es muss demnach zunächst bestimmt werden, welches Antwortverhalten bezüglich des theoretischen Konstruktes Generalisiertes Vertrauen sozialer Erwünschtheit entspricht. Mit anderen Worten: Es muss ermittelt werden, ob es sich bei Generalisiertem Vertrauen um ein für Individuen erstrebenswertes Merkmal handelt oder nicht.

Wird Generalisiertes Vertrauen eher als Ausdruck von Naivität gegenüber anderen Personen angesehen und ist stark mit dem Risiko „ausgenutzt bzw. übervorteilt zu werden“ assoziiert,

³¹ Anm. d. Verf.: Im Folgenden wird die *Trust in People-Scale* auch als Standard-Skala bezeichnet.

³² „Bejahe ich angesichts der Norm, ‚meine ganz persönliche Meinung schonungslos wiederzugeben‘, das Item ‚Ich bin oft ein wenig leichtsinnig‘ bei Zusicherung der Anonymität meiner Aussagen, so werde ich es beispielsweise dann verneinen, wenn ich erwarte, dass meine Antworten von verschiedenen Personen gelesen werden – ich folge dann der sozialen Norm, ‚sich nicht unnötig schlechter zu machen, als man ist‘ oder ‚sich so darzustellen, wie man unter günstigen und nicht unter ungünstigen Bedingungen ist.‘“ (Mummendey/Grau (2014), S. 168)

5. Hypothesen

wird es von Individuen wohl als wenig erstrebenswerte Eigenschaft wahrgenommen. Entsprechend würden Skepsis und Misstrauen gegenüber anderen Menschen als allgemeinverbindliche Norm angenommen werden. Ziel der Probanden dürfte es in diesem Fall sein, sich als Person mit tendenziell gering ausgeprägtem Vertrauen darzustellen bzw. einen möglichst niedrigen Vertrauenslevel zu erreichen. Sozialerwünschtes Antwortverhalten würde somit dem Versuch, möglichst niedrige Testwerte zu erreichen, entsprechen.

Tatsächlich stellt Generalisiertes Vertrauen allerdings eine der wichtigsten gesellschaftlichen Ressourcen dar und entfaltet sowohl auf der sozialen als auch auf der politischen und ökonomischen Ebene Wirkung. Im Vordergrund steht dabei seine Rolle als sogenanntes *social lubricant*, dass die Grundlage für die Kooperation von Individuen bildet.

„The main theoretical underpinning of many accounts of the beneficial effects of generalized trust at the societal level is the claim that generalized trust is the bedrock of cooperation, and that trust-based cooperation is central to the well-being of society. Trust helps avoid or solve collective action problems and reduces transaction costs, for instance in principal-agent situations. Both collective action and principal-agent problems abound in social, political, and economic life and both impair cooperation, so if generalized trust is the cure for these ailments, its beneficial impacts can be expected to be large.“ (Nannestad (2008), S. 428)

Moderne, stark ausdifferenzierte Gesellschaften, zeichnen sich durch eine große Vielfalt von Lebensstilen und sozialen Rollen, ethnische und kulturelle Durchmischung, geografische Mobilität sowie einer Mannigfaltigkeit von Einstellungen und Standpunkten aus. Aus diesen Merkmalen resultiert eine steigende Komplexität sozialer Interaktionsbeziehungen, die nicht zuletzt darauf zurückzuführen ist, dass insbesondere die Notwendigkeit der Kooperation mit Personen, die uns unbekannt sind, an Stellenwert gewinnt. Generalisiertes Vertrauen reduziert an dieser Stelle Opportunitäts-, Transaktions- und Überwachungskosten und erhöht auf diese Weise die Risikobereitschaft von Individuen, was für die Mikroebene zur Folge hat, dass Individuen eher bereit sind, in gemeinsame Unternehmungen mit anderen Personen einzutreten, wobei die Notwendigkeit persönlicher Bekanntschaft in den Hintergrund tritt. Vielmehr erweitert sich die Kooperationsbereitschaft über bloße *face-to-face*-Interaktion hinweg und schließt nun auch fremde Personen mit ein. (Kunz (2004), S. 202; Stolle (2002), S. 397ff)

„Generalized trust is exactly the form of trust which is needed to make the social interaction in complex diversified societies work. It emerges from the multiplicity of impersonal and contractual interactions based on individual autonomy, independence, emancipation, and self-confidence.“ (Stolle (2002), S. 399)

Personen, die ein ausgeprägtes Generalisiertes Vertrauen aufweisen, „do not only engage in mutually beneficial relations more frequently, they are also generally more socially active, engaged, tolerant, and more inclined to support liberal rights, such as minority rights and free speech“ (Stolle (2002), S. 399).

Auf der Makroebene fungiert Generalisiertes Vertrauen „als ein informelles soziales Koordinationsmedium mit großem Steuerungspotential (...), das gesellschaftliche Entwicklung und Integration ermöglicht, wenn gewöhnliche Steuerungsmedien wie Wissen, Macht oder Geld an ihre Grenzen stoßen“ (Kunz (2004), S. 202). Daneben hat es einen positiven Effekt auf demokratische Einstellungen und Verhaltensweisen und fördert auf diese Weise die Stabilität von Demokratien. Die kooperationsbegünstigenden Eigenschaften Generalisierten Vertrauens spielen zusätzlich eine große Rolle für das wirtschaftliche Wachstum, indem sie *joint-ventures* ermöglichen, was die Wahrscheinlichkeit von Innovationen erhöht. Auch auf das Funktionieren

5. Hypothesen

von Märkten hat Generalisiertes Vertrauen einen Einfluss, indem es bei wirtschaftlichen Transaktionen die Verlässlichkeit des Tauschpartners antizipiert. Außerdem unterstützt Generalisiertes Vertrauen soziale Integration in diversifizierten Gesellschaften, indem es die Unterschiede zwischen Menschen überbrückt und wechselseitige Toleranz begünstigt. (Delhey/Newton (2004), S. 152; Kunz (2004), S. 202; Stolle (2002), S. 398f;)

„Clearly, generalized trust is an advantage to people and societies that possess it, as trusters are more likely to initiate cooperative relations that might be beneficial for themselves as well as for their social environment, which benefits from cooperation.“ (Stolle (2002), S. 399)

Bei Generalisiertem Vertrauen scheint es sich somit vornehmlich um eine für Individuen vorteilhafte Eigenschaft zu handeln. Es kann daher angenommen werden, dass Vertrauen gegenüber anderen Menschen als eine allgemeinverbindliche Norm angesehen wird. Absicht der Probanden sollte es demnach sein, sich selbst als besonders vertrauensvoll darzustellen bzw. einen möglichst hohen Vertrauenslevel zu erreichen. Das bedeutet: Sozialerwünschtes Antwortverhalten ist, bezogen auf das theoretische Konstrukt Generalisiertes Vertrauen, vornehmlich mit dem Bemühen gleichzusetzen, einen möglichst hohen Testwert zu erzielen. Es entspricht damit der klassischen Interpretation von „fake-good“.

Ausgehend von der Annahme, dass die Ergebnisse Objektiver Tests bzw. der Hybridklasse aus direkten und indirekten Messverfahren weniger leicht verfälscht werden können als fragebogengestützte Selbstbeurteilungsmaße (Ortner/Proyer (2015), S. 141; Ortner/Schmitt (2014), S. 163), sollte sich für die Messwerte von Objektiv- und Standard-Messung im Allgemeinen die folgenden Beziehungen ergeben:

- H4: Die gemittelten Skalensummenwerte der Objektiv-Messung unterscheiden sich signifikant von den gemittelten Skalensummenwerten der Standard-Messung. Mit anderen Worten: Auf den Objektiv-Skalen wird im Durchschnitt ein anderer Vertrauenslevel angegeben als auf der Standard-Skala.*
- H4a: Bei Anwendung der Standard-Messung ergeben sich im Durchschnitt signifikant höhere Messwerte als bei Anwendung der Objektiv-Messung. Das bedeutet: Auf der Standard-Skala wird im Mittel ein höherer Vertrauenslevel erreicht als auf den Objektiv-Skalen.*
- H5: Die Messergebnisse der Objektiv-Messung korrelieren weniger stark mit der SD-Skala (hier: BIDR-Kurzskala) als die Messergebnisse der Standard-Messung. Soll heißen, dass die Messung Generalisierten Vertrauens per Objektiv-Skalen weniger stark durch sozialerwünschtes Antwortverhalten verzerrt ist als die Messung Generalisierten Vertrauens per Standard-Skala.*
- H6: Die Differenz zwischen den Messwerten der Objektiv-Messung und der Standard-Messung korreliert mit der SD-Skala³³ (hier: BIDR-Kurzskala) und wird somit durch*

³³ Um den Einfluss sozialer Erwünschtheit auf das Antwortverhalten auf den verschiedenen Vertrauensskalen (Objektiv-Skalen und Standard-Skala) zu bestimmen, wird eine sogenannte Kontroll-, Lügen- oder SD-Skala (Social Desirability-Scale) verwendet. Solche Skalen sollen das Ausmaß von Verfälschungen messen und setzen sich aus Items zusammen, „die besonders sensibel auf Tendenzen zu sozial erwünschtem Antworten reagieren“ (Bortz/Döring (2016), S. 439). Im Rahmen der Hauptstudie wird hierzu eine gekürzte Version des *Balanced Inventory of Desirable Responding (BIDR)* von PAULHUS benutzt. (Paulhus (1984); Robinson/Shaver/Wrightsmann (1991);

sozialerwünschtes Antwortverhalten verursacht. Das heißt: Der Unterschied zwischen den gemittelten Skalensummenwerten der Objektiv-Skalen und der Standard-Skala ist auf den Einfluss reaktiver Effekte und dabei im Speziellen auf soziale Erwünschtheit zurückzuführen.

Hintergrund der an dieser Stelle aufgestellten Hypothesen ist die Annahme, dass die Probanden ihr wahres Generalisiertes Vertrauen auf der Standard-Skala überschätzen, während sie auf den Objektiv-Skalen vergleichsweise aufrichtiger antworten. Entsprechend sollte der Einfluss von Verfälschungstendenzen (hier: soziale Erwünschtheit) auf das Antwortverhalten auf den Objektiv-Skalen geringer ausfallen als auf das Antwortverhalten auf der Standard-Skala und der Unterschied zwischen den Messwerten der Skalentypen sollte zumindest in Teilen auf die unterschiedlich starke Anfälligkeit der Skalen für sozialerwünschtes Antwortverhalten zurückgehen. Mit anderen Worten: Die Abweichungen zwischen den Messwerten der Skalen sollten teilweise oder vollständig durch Effekte sozialerwünschten Antwortens bedingt sein.

5.4.1.1 Relevanz der Augenscheinvalidität

Bei der Vermeidung von Messverfälschungen infolge reaktiver Effekte scheint insbesondere die erfolgreiche Verschleierung der Messintention bzw. die geringe Augenscheinvalidität der Objektiv-Items eine große Rolle zu spielen. So konnten verschiedene Studien³⁴ zeigen, dass die Intransparenz der Zieldimension bei Objektiven Tests Verfälschungsabsichten seitens der Probanden nachhaltig verhindern kann. So ziehen ELLIOT/LAWTY-JONES/JACKSON aus den Ergebnissen des Vergleichs zweier Objektiver Tests mit zwei Selbstbeurteilungsfragebogen u.a. folgende Schlussfolgerung:

„Results of this study showed that the lack of face validity was in fact an advantage to the objective tests because it ensured that subjects were unable to fake their answers. In contrast, the high face validity of the self-report questionnaires simply enabled subject dissimulation.“ (Elliot/Lawty-Jones/Jackson (1996), S. 341)

Und ZIEGLER/SCHMIDT-ATZERT/BÜHNER/KRUMM finden bei der Gegenüberstellung eines Objektiven Tests, eines semi-projektiven Messverfahrens und eines Selbstbeurteilungsmaßes u.a. folgendes heraus:

„*Faking the Objective Test*. The most obvious reason for these results is that people were able to recognize what was being measured. Thus, they intentionally distorted their responses. (...) Interestingly, the subjects in this study admitted that faking good in the OLMT³⁵ was harder than in both other tests. Yet, the mean rating of faking success was still average. In other words, people were aware that somehow the objective measure made faking harder, yet they still believed they did a good job. The reason for this probably is the unknown computation of the test score.“ (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 302)

„*Faking the Self-Report*. (...) Obviously, the test intention and score computation are easy to guess and thus, the test can be faked easily.“ (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 303)

Winkler/Kroh/Spiess (2006)) Eine Diskussion dieser Skala erfolgt im Abschnitt „7.1.1.4 Operationalisierung“ des Kapitels „7. Hauptstudie“.

³⁴ Eine Zusammenfassung der Studien erfolgte im Kapitel „2.2.2.5.4 Zur Verfälschungssicherheit von Objektiven Tests“

³⁵ Anm. d. Verf.: *Objective Achievement Motivation Test (OLMT)* entspricht dem in der Studie von ZIEGLER/SCHMIDT-ATZERT/BÜHNER/KRUMM eingesetzten Objektiven Test.

5. Hypothesen

„*Faking the Semi-Projective Test.* (...) However, the fake bad group rated their faking performance regarding the MMG worst but still in an average area. Thus, the test's intention seems to be harder to guess. Nevertheless, faking was possible in both groups³⁶. Unfortunately, the cut-off scores did not have sufficient sensitivity. Thus, faking on the MMG was possible, but the shift in group means was smaller than observed with the self-report measure.“ (Ziegler/Schmidt-Atzert/Bühner/Krumm (2007), S. 303)

Bei der Untersuchung des Einflusses von Verfälschungstendenzen auf die Dimensionalität und die Durchschnittswerte von Objektiven Tests entdecken ARENDASY/SOMMER/HERLE/SCHÜTZHOFFER/INWANSCHITZ schließlich:

„In addition, the results indicated that the fake-job instruction, which was less clear with regard to the level of the latent trait considered ideal hindered respondents' faking attempt. However, the more clear and easier to implement fake-job instruction enabled successful faking.“ (Arendasy/Sommer/Herle/Schützhofer/Inwanschitz (2011), S. 21)

Aus den Studienergebnissen lässt sich ableiten, dass eine erfolgreiche Maskierung der Messintention oder zumindest eine Ambivalenz der *demand characteristics* zielgerichtete Verfälschungsbemühungen wirksam unterbinden können. Letztere Maßnahme bedient sich dabei des Mechanismus, dass aus der Uneindeutigkeit dessen, was als vorteilhaftes Verhalten angesehen wird, eine Unbestimmbarkeit sozialerwünschten Antwortverhaltens resultiert. Dies hat zur Folge, dass durch die Probanden keine systematische Verfälschung der Messwerte generiert werden kann.

Um zu prüfen, ob die Verschleierung der Messintention im Rahmen der Hybridklasse einen besonderen Beitrag zur Verminderung von Messwertverzerrungen infolge reaktiver Effekte bzw. der Vermeidung sozialerwünschten Antwortverhaltens leistet, sollen daher auch die folgenden Hypothesen getestet werden. Hierbei ist allerdings zu beachten, dass die Messintention der Standard-Skala generell transparent gestaltet ist. Das bedeutet: Die Zieldimension der Standard-Skala war, unabhängig von der Identifikation der Messintention der Objektiv-Skalen, für die Probanden unmittelbar einsichtig. Oder mit anderen Worten: Im Gegensatz zu den Items der Objektiv-Skalen verfügten die Items der Standard-Skala grundsätzlich über eine mit der Zieldimension „Generalisiertes Vertrauen“ übereinstimmende Augenscheinvalidität.

5.4.1.1.1 Relevanz der Augenscheinvalidität für die Messwerte von Standard- und Objektiv-Messung

H7a: Unter der Bedingung, dass die Messintention der Objektiv-Messung identifiziert wurde, ergeben sich zwischen den Messwerten von Standard- und Objektiv-Messung keine signifikanten Unterschiede.

Das heißt: Den Probanden war sowohl die Zieldimension der Objektiv-Skalen als auch die Zieldimension der Standard-Skala bekannt. Die Folge ist ein konsistentes Antwortverhalten auf

³⁶ Anm. d. Verf.: Im Rahmen der Studie von ZIEGLER/SCHMIDT-ATZERT/BÜHNER/KRUMM wurde mit sogenannten *Faking*-Instruktionen gearbeitet. Das bedeutet: Es existierten insgesamt zwei Versuchsgruppen und eine Kontrollgruppe. Die Probanden in den Versuchsgruppen bearbeiteten die verschiedenen Messmethoden entweder unter einer *faking-good*-Bedingung, d.h. sie sollten sich möglichst positiv darstellen, oder unter einer *faking-bad*-Bedingung, d.h. sie sollten sich möglichst negativ darstellen. Die Kontrollgruppe erhielt keine Instruktionen hinsichtlich des Antwortverhaltens.

5. Hypothesen

den verschiedenen Vertrauensskalen bzw. eine Angleichung der Messwerte der Vertrauensskalen. Dieses Ergebnis ist vermutlich darauf zurückzuführen, dass beide Skalentypen von einer systematischen und zielgerichteten Verfälschung des Antwortverhaltens betroffen sind.

H7b: Unter der Bedingung, dass die Messintention der Objektiv-Messung nicht identifiziert wurde, fallen die Messwerte der Objektiv-Messung signifikant niedriger aus als die Messwerte der Standard-Messung.

Das heißt: Den Probanden war nur die Zieldimension der Standard-Skala bekannt, während die Zieldimension der Objektiv-Skalen erfolgreich verschleiert werden konnte. Die Folge ist ein inkonsistentes Antwortverhalten auf den verschiedenen Vertrauensskalen. Dieses Ergebnis ist vermutlich darauf zurückzuführen, dass nur auf der Standard-Skala eine systematische und zielgerichtete Verfälschung des Antwortverhaltens möglich war. Es wird angenommen, dass die Probanden ihren Vertrauenslevel auf der Standard-Skala überschätzen, während sie auf den Objektiv-Skalen tendenziell aufrichtiger antworten.

H8a: Unter der Bedingung, dass die Messintention der Objektiv-Messung nicht identifiziert wurde, korrelieren die Messwerte der Objektiv-Messung weniger stark mit der SD-Skala (hier: BIDR-Kurzskala) als unter der Bedingung, dass die Messintention der Objektiv-Messung identifiziert wurde.

Mit anderen Worten: Wenn die Verschleierung der Messintention der Objektiv-Skalen erfolgreich verläuft, dann sollten die Messwerte der Objektiv-Skalen weniger stark durch sozialerwünschtes Antwortverhalten verzerrt sein.

H8b: Die Korrelation der Messwerte der Standard-Messung mit der SD-Skala (hier: BIDR-Kurzskala) bleibt unabhängig von der Identifikation der Messintention der Objektiv-Messung konstant.

Da die Items der Standard-Skala prinzipiell über eine mit der Zieldimension übereinstimmende Augenscheinvalidität verfügen bzw. den Probanden unmittelbar Einblick in die Messintention der Standard-Skala gewähren, sollten die Messwerte der Standard-Skala, unabhängig von der erfolgreichen Verschleierung der Messintention der Objektiv-Skalen, ähnlich stark durch sozialerwünschtes Antwortverhalten verfälscht sein.

H9a: Unter der Bedingung, dass die Messintention der Objektiv-Messung identifiziert wurde, korrelieren die Messwerte der Objektiv- und Standard-Messung annähernd gleich stark mit der SD-Skala (hier: BIDR-Kurzskala).

Gelingt die Verschleierung der Messintention der Objektiv-Skalen nicht bzw. ist die Zieldimension beider Skalentypen für die Probanden einsichtig, so sollten die Messwerte beider Skalentypen gleichermaßen durch sozialerwünschtes Antwortverhalten verfälscht sein.

5. Hypothesen

H9b: Unter der Bedingung, dass die Messintention der Objektiv-Messung nicht identifiziert wurde, korrelieren die Messwerte der Objektiv-Messung weniger stark mit der SD-Skala (hier: BIDR-Kurzskala) als die Messwerte der Standard-Messung.

Das bedeutet: Wenn die Verschleierung der Messintention der Objektiv-Skalen gelingt, dann sollten die Messwerte der Objektiv-Skalen weniger stark durch sozialerwünschtes Antwortverhalten verfälscht sein als die Messwerte der Standard-Skala, bei denen die Zieldimension offen lag.

5.4.1.1.2 Relevanz der Augenscheinvalidität für die Differenz zwischen Standard- und Objektiv-Messung

H10: Die Differenz zwischen den Messwerten der Objektiv-Messung und der Standard-Messung korreliert unabhängig von der Identifikation der Messintention der Objektiv-Messung mit der SD-Skala (hier: BIDR-Kurzskala)

Die Abweichungen zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala kommen unabhängig davon, ob die Zieldimension beider Skalentypen für die Probanden einsichtig war oder ausschließlich die Zieldimension der Standard-Skala für die Probanden offen lag, zumindest in Teilen durch den Einfluss sozialerwünschten Antwortverhaltens zustande.

H10a: Unter der Bedingung, dass die Messintention der Objektiv-Messung identifiziert wurde, korreliert die Differenz zwischen den Messwerten der Objektiv-Messung und der Standard-Messung sowohl negativ als auch positiv mit der SD-Skala (hier: BIDR-Kurzskala)

Ist sowohl die Messintention der Objektiv-Skalen als auch die Messintention der Standard-Skala transparent bzw. die Zieldimensionen beider Skalentypen konnten von den Probanden bestimmt werden, so lässt sich keine klare Aussage darüber treffen, welche Auswirkungen sozialerwünschtes Antwortverhalten auf den Messwertunterschied zwischen den Skalen hat. Auf der einen Seite könnte ein auf beiden Skalentypen konsistentes Antwortverhalten die Folge sein. Entsprechend der Präsentationsreihenfolge der Skalen gleichen die Probanden möglicherweise ihre Antworten auf den Skalen aneinander an, was einem negativen Zusammenhang von Messwertdifferenz und SD-Skala entsprechen würde. Der Abstand zwischen den Messwerten der unterschiedlichen Skalentypen verringert sich somit infolge sozialerwünschten Antwortverhaltens.

Auf der anderen Seite könnte die Probanden auf den verschiedenen Skalentypen inkonsistent Antworten. Auch hierbei dürfte die Präsentationsreihenfolge der Skalentypen eine Rolle spielen. In jedem Fall würde ein solches Antwortverhalten einem positiven Zusammenhang von Messwertdifferenz und SD-Skala entsprechen, sodass der Abstand zwischen den Messwerten der unterschiedlichen Skalentypen infolge sozialerwünschten Antwortverhaltens vergrößert wird. Allerdings kann nicht vorhergesagt werden, ob die Messwertdifferenz durch ein *fake good* oder *fake bad* bzw. eine Korrektur des Vertrauensniveaus auf einer Skala nach oben oder nach unten in Relation zur jeweils anderen Skala hervorgerufen wird.

H10b: Unter der Bedingung, dass die Messintention der Objektiv-Messung nicht identifiziert wurde, korreliert die Differenz zwischen den Messwerten der Objektiv-Messung und der Standard-Messung positiv mit der SD-Skala (hier: BIDR-Kurzskala)

Wurde die Messintention der Objektiv-Skalen erfolgreich verschleiert bzw. ist nur die Zieldimension der Standard-Skala für die Probanden einsehbar, sollte sozialerwünschtes Antwortverhalten eine Vergrößerung des Abstandes zwischen den Messwerten der unterschiedlichen Skalentypen bewirken. Mit anderen Worten: Infolge der Intransparenz der Zieldimension der Objektiv-Skalen sollten die Probanden zu inkonsistentem Antwortverhalten neigen.

H10c: Unter der Bedingung, dass die Messintention der Objektiv-Skalen identifiziert wurde, ist die Korrelation der Differenz zwischen Messwerten der Objektiv-Messung und der Standard-Messung mit der SD-Skala (hier: BIDR-Skala) vornehmlich darauf zurückzuführen, dass die Messwerte beider Skalentypen gleichermaßen durch sozialerwünschtes Antwortverhalten verfälscht werden.

Die Angleichung bzw. das Auseinanderdriften der Messwerte der Objektiv-Skalen und der Standard-Skala infolge sozialerwünschten Antwortverhaltens, wenn die Zieldimension beider Skalentypen offenliegt, ist sowohl auf Verzerrungen der Messwerte der Objektiv-Skalen als auch auf Verzerrungen der Messwerte der Standard-Skala zurückzuführen. Hypothese 10c stellt in diesem Sinne eine Verbindung von Hypothese 9a und Hypothese 10a dar.

H10d: Unter der Bedingung, dass die Messintention der Objektiv-Messung nicht identifiziert wurde, ist die Korrelation der Differenz zwischen den Messwerten der Objektiv-Messung und der Standard-Messung mit der SD-Skala (hier: BIDR-Skala) vornehmlich auf die Verzerrung der Messwerte der Standard-Skala durch sozialerwünschtes Antwortverhalten zurückzuführen.

Die Abweichung der Messwerte der Objektiv-Skalen von den Messwerten der Standard-Skala infolge sozialerwünschten Antwortverhaltens, wenn die Zieldimension der Objektiv-Skalen verschleiert werden konnte, ist insbesondere auf Verzerrungen der Messwerte der Standard-Skala zurückzuführen. Hypothese 10d stellt in diesem Sinne eine Verbindung von Hypothese 9b und Hypothese 10b dar.

Grundannahme der hier aufgestellten Hypothesen ist, dass insbesondere die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen bzw. die Tatsache, dass sich die Objektiv-Skalen aus Items zusammensetzen, die eine geringe Augenscheinvalidität bezüglich der Zieldimension „Generalisiertes Vertrauen“ aufweisen, die Anfälligkeit für Verfälschungen im Vergleich zur Standard-Skala reduziert. Entsprechend sollten sich nur dann überzufällige Unterschiede zwischen den Messwerten der verschiedenen Skalentypen ergeben, wenn die Maskierung der Zieldimension der Objektiv-Skalen gelingt, wobei die Probanden ihren wahren Vertrauenslevel auf der Standard-Skala überschätzen und auf den Objektiv-Skalen tendenziell aufrichtiger antworten. Hierzu sollte unter der Bedingung, dass die Verschleierung der Messintention gelingt, die Anfälligkeit der Objektiv-Skalen für sozialerwünschtes Antwortverhalten geringer ausfallen als die Anfälligkeit der Standard-Skala für

5. Hypothesen

sozialerwünschtes Antwortverhalten. Und unter der Bedingung, dass die Messintention der Objektiv-Skalen identifiziert wurde, sollten die verschiedenen Skalentypen eine ähnlich starke Anfälligkeit für sozialerwünschtes Antwortverhalten aufweisen.

Der Unterschied zwischen den Messwerten der verschiedenen Skalen sollte somit in jedem Fall auf den Einfluss sozialer Erwünschtheit zurückzuführen sein. Allerdings sollte sich sozialerwünschtes Antwortverhalten je nachdem, ob die Messintention beider Skalentypen oder nur eines der Skalentypen offen liegt, unterschiedlich auswirken.

Sind sowohl die Messintention der Objektiv-Skalen als auch der Standard-Skala für die Probanden einsichtig, so kann sozialerwünschtes Antwortverhalten sowohl eine Angleichung als auch eine Abweichung der Messwerte und dem zur Folge eine Verringerung oder Vergrößerung der Differenz zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala bewirken. Es wird dabei angenommen, dass durch die Transparenz der Zieldimension eine systematische und zielgerichtete Verfälschung der Messwerte beider Skalentypen ermöglicht wird. Hieraus resultiert dann entweder ein konsistentes oder inkonsistentes Antwortverhalten auf den verschiedenen Vertrauensskalen.

Ist nur die Messintention der Standard-Skala für die Probanden ersichtlich, sollte sozialerwünschtes Antwortverhalten eher eine Vergrößerung der Differenz zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala bewirken. In diesem Fall sollte insbesondere die Standard-Skala für systematische Verfälschungen anfällig sein, während der Effekt sozialerwünschten Antwortverhaltens auf die Messwerte der Objektiv-Skalen geringer ausfällt. Die Folge ist ein inkonsistentes Antwortverhalten, das vornehmlich auf Verzerrungen der Messwerte der Standard-Skala zurückzuführen ist. Mit anderen Worten: Der Unterschied zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala wird durch die Anfälligkeit der Standard-Skala für sozialerwünschtes Antwortverhalten verursacht.

Sollten die Objektiv-Skalen unter der Bedingung, dass ihre Messintention nicht identifiziert werden konnte, weniger stark durch soziale Erwünschtheit verfälscht sein als unter der Bedingung, dass ihre Messintention identifiziert werden konnte, stellt dies ebenfalls einen Hinweis auf die Bedeutsamkeit geringer Augenscheinvalidität für die Reduktion von Verfälschungseffekten dar.

5.5 Radiusproblem

Wie bereits an anderer Stelle³⁷ detailliert beschrieben, stellt das Radiusproblem eine der Hauptproblematiken der Vertrauensforschung dar und beschreibt dabei im Wesentlichen eine mögliche Bedrohung der Validität von Vertrauensmessungen. Im Kern wird der Frage nachgegangen, welche Personen die Probanden als Vertrauensreferenz heranziehen, wenn innerhalb der Items der standardmäßig zur Erfassung von Generalisiertem Vertrauen eingesetzten *Trust in People-Scale* (hier: Standard-Skala) von „most people“ die Rede ist. Dies hat den Hintergrund, dass die Art des abgefragten Vertrauens vom Personenkreis abhängig ist, auf den sich das Vertrauen bezieht. Entsprechend wird Partikulares Vertrauen abgefragt, wenn die Probanden die *ingroup*, d.h. Personen, die ihnen unmittelbar persönlich bekannt sind oder aus ihrem näheren sozialen Umfeld stammen, als Bezugsrahmen ihres Vertrauens angeben. Generalisiertes Vertrauen wird dagegen abgefragt, wenn die Probanden die *outgroup*, d.h.

³⁷ Vgl. Abschnitt „4.3.2 Radius generalisierten Vertrauens“

5. Hypothesen

Fremde oder Personen, hinsichtlich derer man nur wenige Informationen besitzt, als Bezugsgruppe ihres Vertrauens nennen. (Bjørnskov (2006), S. 2; Delhey/Newton/Welzel (2011), S. 787ff; Fukuyama (2000), S. 4f; Fukuyama (2001), S. 8f; Glaeser/Laibson/Scheinkam/Soutter (2000), S. 815; Reeskens/Hooghe (2008), S. 516)

Das Radiusproblem stellt sich nun wie folgt dar: Ziel der *Trust in People-Scale* ist die Abfrage Generalisierten Vertrauens, sodass die Probanden vornehmlich generalisierte Andere als Referenzgruppe ihres Vertrauens angeben sollten. Werden nun aber in der Mehrzahl bzw. einer Vielzahl der Fälle spezifische Andere als Bezugspunkt des Vertrauens angegeben, so wird in Wahrheit entweder ein von der intendierten Zieldimension abweichendes Konstrukt, nämlich „Partikulares Vertrauen“, gemessen oder es findet eine Diffusion der verschiedenen Formen des Vertrauens statt. In beiden Fällen könnten Messwertverzerrungen die Folge sein. So konnten STURGIS/SMITH Folgendes zeigen:

„(...) respondents who report having thought about people with whom they are personally familiar when formulating their answer are substantially more trusting than those who do not. Similarly, respondents for whom the questions evoked abstract categories such as „people in general“, „the general public“, or „everyone“ report significantly lower levels of trust than respondents who did not provide this type of verbatim response.“ (Sturgis/Smith (2010), S. 83)

Auch DELHEY/NEWTON/WELZEL berichten:

„Research shows that the further people move from their immediate circle of friends, colleagues, and neighbors, the less likely they are to trust. (...) Differences in trust levels can be interpreted sensibly only when trust radiuses are similar.“ (Delhey/Newton/Welzel (2011), S. 789)

Dabei berufen sich DELHEY/NEWTON/WELZEL vornehmlich auf Studienergebnisse verschiedener anderer Forscher. So heißt es bei WELCH/SIKKING/LOVELAND bezüglich der USA:

„Individuals are clearly more likely to trust those with whom they are acquainted than strangers, and this trust in acquaintances seems to be critical for understanding how much persons are willing to trust those they don't know.“ (Welch/Sikking/Loveland (2007), S. 38)

Und SCHUPP/WAGNER finden für Deutschland heraus:

„Nicht überraschend ist, dass 86% der Deutschen ihrer eigenen Familie „sehr viel“ vertrauen, während Freunden lediglich in knapp 50% in gleichem Maße vertraut wird. Immerhin wird zusammengekommen zu 93% Freunden „ziemlich viel“ und „sehr viel“ Vertrauen geschenkt. Nahezu drei Viertel aller Deutschen vertrauen auch ihren Nachbarn entsprechend. Dabei ist das Vertrauen zu Nachbarn in Westdeutschland mit 72% häufiger vorhanden als in Ostdeutschland (61%). Fremden, denen das erste Mal begegnet wird, bringen die Bürger in Deutschland erwartungsgemäß nur in geringem Maß Vertrauen entgegen. Lediglich jeder achte Erwachsene hat hier Vertrauen; ein Viertel hat dagegen geäußert, überhaupt kein Vertrauen aufzubringen.“ (Wagner/Schupp (2004), S. 312)

Ausgehend von diesen Studienergebnissen, könnte eine Durchmischung von Partikularem und Generalisiertem Vertrauen somit, je nach anvisierter Zieldimension, eine Überschätzung des Ausmaßes Generalisierten Vertrauens oder eine Unterschätzung des Ausmaßes Partikularen Vertrauens nach sich ziehen. Daneben würden unberücksichtigte Interpretationsunterschiede hinsichtlich der Phrase „most people“ die Vergleichbarkeit der Messergebnisse einschränken. (Delhey/Newton/Welzel (2011), S. 787ff; Glaeser/Laibson/Scheinkam/Soutter (2000), S. 815; Reeskens/Hooghe (2008), S. 516)

5. Hypothesen

Im Rahmen der Hauptstudie soll nun überprüft werden, ob die Verwendung der in der Pilotstudie entwickelten Vertrauensskalen zur Lösung des Radiusproblems beitragen kann. Um dies zu bestimmen, sollen die folgenden Hypothesen überprüft werden:

H11a: Bei der Messung von Vertrauen mit Hilfe der Objektiv-Skalen werden in der Mehrzahl der Fälle die outgroup bzw. generalisierte Andere als Vertrauensreferenz angegeben.

H11b: Bei der Messung von Vertrauen mit Hilfe der Objektiv-Skalen werden häufiger die outgroup bzw. generalisierte Andere als Vertrauensreferenz angegeben als bei der Vertrauensmessung mit Hilfe der Standard-Skala.

Als Erklärung für einen möglichen Effekt könnten dabei insbesondere die Formulierung und die inhaltliche Schwerpunktsetzung der Objektiv-Items eine Rolle spielen.

5.5.1 Radiusproblem und soziale Erwünschtheit

Das Radiusproblem hat dabei strenggenommen auch Einfluss auf die Überprüfung der im vorangegangenen Abschnitt aufgestellten Hypothesen zum Effekt sozialerwünschten Antwortverhaltens. So sind potentielle Messwertunterschiede zwischen den Skalen möglicherweise darauf zurückzuführen, dass die Standard-Skala erwiesenermaßen nicht exklusiv Generalisiertes Vertrauen abbildet, sondern zumindest teilweise auch Partikulares Vertrauen erfasst wird. Ebenso könnten beide Skalentypen nicht dazu in der Lage sein, ausschließlich Generalisiertes Vertrauen zu messen. Ähnliches gilt für die Beeinflussbarkeit durch sozialerwünschtes Antwortverhalten. So neigen Probanden, die generalisierte Andere als Vertrauensreferenz angegeben haben, und Probanden, die spezifische Andere als Vertrauensreferenz nennen, möglicherweise in unterschiedlichem Umfang zu sozialerwünschtem Antworten. Unterschiede hinsichtlich der Anfälligkeit der Skalentypen für reaktive Effekte könnten in diesem Fall durch die Diffusion der Zieldimensionen Generalisiertes und Partikulares Vertrauen moderiert sein.

Es werden dementsprechend die Hypothesen 7a bis 10d zusätzlich auch unter der Bedingung getestet, dass die Probanden die *outgroup* als Bezugsrahmen ihres Vertrauens angegeben haben.

6. Pilotstudie

6.1 Methodik

6.1.1 Forschungsdesign

6.1.1.1 Erhebungskonzept

Das Erhebungskonzept der Pilotstudie kann im Wesentlichen als einfache quantitative Querschnittsbefragung im *Ex-post-facto*-Design beschrieben werden. Es findet demnach eine einzelne Befragung zu einem bestimmten Zeitpunkt statt. Strenggenommen verteilt sich diese einmalige Befragung allerdings auf mehrere Erhebungszeitpunkte, um ein Maximum an Fällen zu erreichen. (Bernard (2013), S. 100, S. 104f; Diekmann (2011), S. 305f; Schnell/Hill/Esser (2008), S. 212, S. 228f, S. 230ff)

Tab. 6.1: The One-Shot Case Study Design

	<i>Time 1</i>		<i>Time 2</i>	
	Assignment	Pretest	Intervention/Treatment	Posttest
Versuchsgruppe			X	O
Keine Kontrollgruppe				

Ziel der Pilotstudie war die Entwicklung und Testung einer Fragebatterie zur Erfassung Generalisierten Vertrauens, die entsprechend der Designrichtlinien für die Hybridklasse aus direkten und indirekten Messverfahren erstellt wurde. Diese Vorgaben leiten sich im Wesentlichen aus der Konzeption *Objektiver Tests* nach CATTELL ab. *Objektive Tests* beschreiben Messverfahren, die das Verhalten bzw. die Einstellungen eines Individuums (1) in einer standardisierten Situation erfassen, (2) ohne dass der Befragte eine Selbstbeurteilung vollzieht und (3) ohne dass das Messinstrument Hinweise auf die Messintention gibt. (Fisseni (1997), S. 309; Fisseni (2004), S. 210f; Schmidt (1975), S. 19)

Um dem Umstand der standardisierten Situation gerecht zu werden, wurde als Untersuchungsinstrument ein standardisierter Fragebogen verwendet, in den die Fragebatterie eingebettet war.

Die Vermeidung einer Selbstbeurteilung der Befragten wurde durch eine entsprechende Formulierung der verwendeten Items realisiert. So nahmen die jeweiligen Statements vornehmlich Bezug auf Menschen im Allgemeinen und deren Verhalten und Eigenschaften. Außerdem kamen Items zum Einsatz, deren Gegenstand alltägliche Situationen, politische Institutionen sowie Normen und Werte waren. Die Befragten mussten somit weder ihr eigenes Verhalten beurteilen noch eine dedizierte Selbstauskunft geben.

Die Vermeidung eines Hinweises auf die Messintention wurde bewerkstelligt, indem bei der Auswahl und Formulierung der Items auf eine geringe Augenscheinvalidität hinsichtlich der Zieldimension „Generalisiertes Vertrauen“ geachtet wurde. So wurde in den Items kein direkter Bezug auf die Zieldimension genommen, sondern Begleiterscheinungen, Korrelate und Folgen Generalisierten Vertrauens als Indikatoren verwendet, um so Generalisiertes Vertrauen auf indirektem Wege zu erfassen. Diese ergaben sich im Wesentlichen aus den theoretischen Konzepten zum Konstrukt „Generalisiertes Vertrauen“.

6.1.1.2 Stichprobe

Zielsetzung der Pilotstudie war es, eine möglichst große Fallzahl zu generieren, anhand der eine der Hybridklasse entsprechende Skala zur Erfassung Generalisierten Vertrauens entwickelt werden sollte. Gängige Praxis bei der Entwicklung von Skalen ist es sogenannte *convenience sample*, bestehend aus Studierenden³⁸, zu verwenden. (Eid/Gollwitzer/Schmitt (2010), S. 63; Robinson, J.P.; Shaver, P.R., Wrightsman, L.S. (1991))

Solche Gelegenheitsstichproben oder Ad-hoc-Stichproben sind den nicht-probabilistischen Stichprobenverfahren zuzuordnen, die dadurch gekennzeichnet sind, dass „die Auswahl der Untersuchungsobjekte willkürlich und/oder bewusst/absichtsvoll durch den Forschenden“ (Bortz/Döring (2016), S. 305) erfolgt. Im Rahmen eines *convenience sample* „werden willkürlich Personen oder Objekte für die Stichprobe einer Untersuchung ausgewählt, die gerade zur Verfügung stehen oder leicht zugänglich sind“ (Bortz/Döring (2016), S. 306). Die Problematik bei einem solchen Modus der Stichprobenziehung besteht darin, dass keine klare Definition der Zielpopulation abgegeben werden kann und dementsprechend allenfalls eine indirekte Generalisierung der Ergebnisse stattfinden kann. Die Verwendung einer solchen Stichprobe in der vorliegenden Studie kann allerdings trotzdem als opportun und zulässig angesehen werden, da das Forschungsziel keine Deskription der Merkmale einer Population darstellt. Vielmehr handelt es sich um eine Studie mit explorativem Charakter, innerhalb der ein neuartiges Untersuchungsinstrument entwickelt und getestet werden sollte. Zudem war die Realisierung einer probabilistischen Stichprobe aus vornehmlich finanziellen und organisatorischen Gründen nicht möglich. Die Verwendung eines *convenience sample* ist somit durch die Zielsetzung der Pilotstudie gerechtfertigt. (Babbie (1975), S. 167ff; Bernard (2013), S. 162ff; Bortz/Döring (2016), S. 305ff; Diekmann (2011), S. 194f; Fink (1995), S. 17f; Hussy/Jain (2002), S. 126, S. 193; Leiner (2014), S. 4ff)

Im Rahmen der Datenerhebung konnte ein Stichprobenumfang von N=220 Studierenden realisiert werden. Die Stichprobe enthielt 132 (60,3%) weibliche und 87 (39,7%) männliche Studierende, die ebenfalls zwischen 18 und 70 Jahren ($\bar{x} = 22,2752$ Jahre, $\sigma = 5,26158$) alt waren. Insofern waren Frauen in der Stichprobe vergleichsweise überrepräsentiert. Die Obergrenze der Altersrange ergab sich aus einem einzelnen Fall, der stark vom Durchschnittsalter abwich. Das bedeutet hinsichtlich des Alters war die Stichprobenzusammensetzung bis auf diesen Ausreißer sehr homogen.

Tab. 6.2: Zusammensetzung der Stichprobe (Pilotstudie)

Stichprobe	Geschlecht				Alter (Jahre)				
	N	♀	♂	Keine Angabe	μ	σ	Min	Max	Keine Angabe
Netto-Stichprobe	220	132	87	1	22,28	5,26158	18	70	2

³⁸ Tatsächlich bilden Personen aus sogenannten „WEIRD societies“ (Western, Educated, Industrialized, Rich and Democratic), und im Speziellen amerikanische Studierende, die Datengrundlage für einen Großteil der empirischen Untersuchungen in den Sozial- und Wirtschaftswissenschaften. Erkenntnisse innerhalb dieser Disziplinen basieren somit vornehmlich auf einer Subpopulation, für die im Rahmen einer Speziesgeneralisierung Repräsentativität für die gesamte Menschheit angenommen wird. Es ist daher fraglich, ob Ergebnisse auf Basis von WEIRD-Stichproben auf Populationen übertragbar sind, die außerhalb dieser gesellschaftlichen und sozioökonomischen Rahmung existieren. So kann angenommen werden, dass sich Verzerrungen allein dadurch ergeben, dass sowohl die Forscher als auch die verwendeten Untersuchungsinstrumente aus WEIRD Populationen stammen. Gleichzeitig zeigt sich aber, dass hinsichtlich grundlegender Aspekte von Kognition, Motivation und Verhalten große Gemeinsamkeiten zwischen verschiedenen Bevölkerungsgruppen bestehen. (Henrich/Heine/Norenzayan (2010))

Nach BÜHNER und MENDOZA/STAFFORD/STAUFER ist bereits ab einer Stichprobenmindestgröße von $N = 100$ eine zuverlässige Schätzung der Reliabilität eines Tests möglich. FISSENI gibt eine differenziertere Einschätzung der notwendigen Stichprobengröße ab. Hiernach gilt eine Stichprobe mit $N = 150-300$ als mittelgroß und entsprechend durchschnittlich gut geeignet. Eine Stichprobengröße von $N = 220$ kann daher für die Prüfung eines Tests im Rahmen der Testentwicklung als angemessen angesehen werden. (Bühner (2006), S. 51, S. 140; Fisseni (1997), S. 124; Mendoza/Stafford/Staufner (2000), S. 367)

6.1.1.3 Feldzugang

Die Präsentation der Fragebatterie zur Erfassung Generalisierten Vertrauens erfolgte im Rahmen eines standardisierten Fragebogens, der Studierenden im Kontext von drei Vorlesungen des Instituts für Politikwissenschaft der Universität Regensburg vorgelegt wurde. Als *Gatekeeper* fungierten hierbei die jeweiligen Lehrenden. Um das Problem des *Unit-Nonresponse* zu minimieren und die Ausbeute an Fällen zu maximieren, sollte die Befragung jeweils zu Beginn der Lehrveranstaltung stattfinden. Die Lehrenden wurden deshalb gebeten, ca. 15-20 Minuten am Anfang der Veranstaltung für die Befragung zur Verfügung zu stellen. Befragungen zu anderen Zeitpunkten hätten die Teilnahmebereitschaft der Studierenden erheblich reduziert.

Um Doppelbefragungen zu vermeiden, wurde jeweils im Vorhinein der Erhebungen darauf hingewiesen, dass nur diejenigen Personen teilnehmen sollten, die nicht bereits an einer vorherigen Befragungsrunde teilgenommen hatten.

6.1.1.4 Operationalisierung

Das im Rahmen der Pilotstudie zum Einsatz kommende Untersuchungsinstrument entspricht einem standardisierten Fragebogen³⁹. Dieser umfasst sechs verschiedene Itembatterien, in denen insgesamt 64 Items im offenen und geschlossenen Frageformat enthalten sind. Erfasst werden hauptsächlich Aspekte Generalisierten Vertrauens und dessen potentielle Korrelate sowie die individuelle Tendenz zu sozialerwünschtem Antwortverhalten. Nachfolgend sollen die einzelnen Itembatterien bzw. Items entsprechend ihrer chronologischen Reihenfolge im Fragebogen vorgestellt werden und die für das jeweilige Konstrukt gewählte Operationalisierung theoretisch untermauert werden.

6.1.1.4.1 Glück und subjektives Wohlbefinden/Zufriedenheit

Die Items v1.1 und v1.2 fragen das subjektive Wohlbefinden und die Lebenszufriedenheit der Probanden ab. Subjektives Wohlbefinden ist insofern mit Generalisiertem Vertrauen verbunden, als dass Menschen, die mit ihrem Leben zufrieden sind, eine eher positive Einstellung gegenüber Fremden aufweisen und eine optimistischere Weltwahrnehmung besitzen. Lebenszufriedenheit bewirkt darüber hinaus, dass man sich selbst positiver sieht und daraus vorteilhafte Zukunftsaussichten ableitet. Außerdem haben glückliche Menschen mehr Gelegenheiten zur Kooperation, da sie auf andere Menschen attraktiver wirken. Aus positiven Erfahrungen mit diesen Interaktionsbeziehungen leitet sich in Konsequenz Generalisiertes

³⁹ Vgl. Anhang IV: Fragebogen, IV.a Pilotstudie

Vertrauen ab. Gleichzeitig führt aber auch die Möglichkeit, dem sozialen Umfeld vertrauen zu können zu einer höheren subjektiven Zufriedenheit. (Bjørnskov (2003), S. 4f, S. 6, S. 10; Bjørnskov (2006), S. 1; Bjørnskov (2008), S. 271; Delhey/Newton (2002), S. 4f, S. 17ff, S. 22; Delhey/Newton (2004), S. 153, S. 163; Helliwell (2003), S. 346f, S. 355; Helliwell/Putnam (2004), S. 1442, S. 1444; Uslaner (2002), S. 81, S. 84f, S. 94)

Die Abfrage des subjektiven Wohlbefindens gliedert sich in zwei Teile. So fragt Item v1.1 das Wohlbefinden zum Erhebungszeitpunkt ab, während sich Item v1.2 auf die subjektive Lebenszufriedenheit im Allgemeinen bezieht. Die jeweils angelegte Antwortskala weist die folgenden Randausprägungen auf: „*Sehr unzufrieden/unglücklich*“ (1) und „*Sehr zufrieden/glücklich*“ (5). Außerdem wurde eine „*Keine-Angabe*“-Kategorie hinzugefügt.

6.1.1.4.2 Generalisiertes Vertrauen

„One reason why trust is such a hard term even to define and why it may have so many apparent meanings in the vernacular an even scholarly work is that it is not a primitive term that is unanalyzable into other terms. Rather, it is essentially a reductive term, in the following sense: Trust is not a primitive, something that we just know by inspection, as the color blue might be primitive, at least for ordinary people who do not think of it as a problem in optics. Rather, it is reducible to other things that go into determining trust. (...) As a reductive term, trust is similar to other major theoretical terms such as power, equality, and justice that must be explained and are not primitive terms but are reducible to other terms.“ (Hardin (2002), S. 56f)

Generalisiertes Vertrauen darf dementsprechend nicht als unidimensionales Konstrukt betrachtet werden, sondern enthält verschiedene Facetten bzw. ist durch eine Vielzahl von Aspekten determiniert und stellt insofern ein multidimensionales Konstrukt dar. Diesem Sachverhalt wurde dadurch Rechnung getragen, dass sich die Operationalisierung Generalisierten Vertrauens an dessen Determinanten orientiert. Auf diese Weise konnte außerdem insbesondere die Vorgabe, nur Items zu verwenden, die keine oder nur eine geringe mit der Zieldimension übereinstimmende Augenscheinvalidität aufweisen, bestmöglich umgesetzt werden.

Neben Eigenentwicklungen enthält die Skala zur Erfassung Generalisierten Vertrauens auch Items aus bereits bestehenden Vertrauensskalen, wie *Interpersonal Trust Scale (ITS)*, *Philosophies of Human Nature (PHN)* und *Machiavellismus (Mach IV bzw. Mach V)*. Voraussetzung für die Selektion der Items war dabei, dass sie den Designvorgaben der Hybridklasse entsprechen.

Zur Bewertung der Items wurde eine 5-stufige Ratingskala verwendet, deren Randausprägungen „*Stimme nicht zu*“ (1) und „*Stimme zu*“ (5) lauteten, ergänzt wurde diese Skala durch eine „*Weiß nicht*“-Kategorie (0). Um die kognitive Aufmerksamkeit der Befragten zu erhöhen bzw. aufrechtzuerhalten, wurde bei der Zusammenstellung der Items auf eine heterogene Polung geachtet. Es wurden daher sowohl negativ als auch positiv formulierte Statements verwendet.

Es sei bereits an dieser Stelle angemerkt, dass die nachfolgend dargestellte Art und Weise der Faktorstruktur aufgrund der im Zuge von Item- und Reliabilitätsanalyse vorgenommenen Eliminierung von Items, nicht in gleicher Form reproduziert werden konnte.

6.1.1.4.2.1 Optimismus

Generalisiertes Vertrauen basiert in hohem Maße auf einer optimistischen Weltanschauung. Die Annahme einer positiven Zukunft und die Überzeugung, dass die Welt ein guter und freundlicher Ort ist, lässt die Umwelt als weniger risikobehaftet erscheinen. Vertrauensvolle Menschen zeichnen sich dementsprechend durch einen stark ausgeprägten Optimismus gegenüber den Motivationen anderer aus. Auf Grundlage dieser positiven Wahrnehmung der menschlichen Natur wird dann bestimmt, ob es opportun und ausreichend gefahrenfrei ist, eine Kooperationsbeziehung einzugehen. Vertrauensvolle Menschen nehmen ihre Umwelt in der Form selektiv wahr, dass ab einem bestimmten Punkt keine weiteren rationalen Beweise mehr für die Bestimmung der Vertrauenswürdigkeit eines Interaktionspartners notwendig sind und ein sogenannter *cognitive leap* stattfindet, der losgelöst von Erfahrung und vernunftgeleiteten Begründungen, den anderen als vertrauenswürdig klassifiziert.

Diese selektive Wahrnehmung gilt gleichermaßen für das Pendant vertrauenswürdiger Menschen. Menschen mit einem niedrigen Generalisierten Vertrauenslevel weisen eine starke Tendenz zu Pessimismus auf und interpretieren Interaktionen mit anderen Personen eher negativ. Sie gehen davon aus, dass man Fremden mit Vorsicht begegnen sollte und dass Menschen nicht immer auch am Wohlergehen des anderen interessiert sind.

(Lewis/Weigert (1985a), S. 970; Rothstein/Uslaner (2005), S. 51; Seligman (1991); Uslaner (2002), S. 76ff, S. 95, S. 109f; Uslaner (2004), S. 502; Uslaner (2005), S. 77; Uslaner (2008a), S. 103, S. 108, S. 113; Uslaner (2008c), S. 733; Uslaner (2013), S. 629ff; Uslaner/Brown (2005), S. 871)

Entsprechend sollten Personen mit ausgeprägtem Generalisiertem Vertrauen eher positive Zukunftsaussichten vertreten sowie davon ausgehen, dass sich schwierige Situationen eher vorteilhaft, denn nachteilig entwickeln. Es ergeben sich daher die folgenden Items zur Erfassung dieses Aspekts Generalisierten Vertrauens:

- *Die Zukunft erscheint vielversprechend.* (Interpersonal Trust Scale)
- *Wenn jemand seinen Geldbeutel auf dem Tisch eines Restaurants vergisst, wird er in der Mehrzahl der Fälle nicht zurückgegeben.* (vgl. Schmidt (1975))
- *Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte.*
- *Auf die Dauer gesehen, gleichen sich die guten und schlechten Dinge im Leben aus.* (I-E-Scale)

6.1.1.4.2.2 Subjektive Schicksalskontrolle/Selbstkontrolle (Locus of Control)

Die Annahme, Kontrolle über das eigene Schicksal zu besitzen, ist eng verbunden mit der bereits beschriebenen optimistischen Weltanschauung. So gehen vertrauensvolle Menschen nicht nur davon aus, dass sie sich innerhalb einer ihnen freundlich gesonnenen Umwelt bewegen, sondern dass sie ebenfalls dazu in der Lage sind, diese Umwelt positiv zu gestalten und zu entwickeln. Die umgebende Welt hält dabei grenzenlose Opportunities bereit.

Dagegen führt die Überzeugung, dass andere Herr über das eigene Schicksal sind, zur Entwicklung von Fatalismus und Misstrauen. Misstrauische Personen nehmen an, dass Fremde Konkurrenten hinsichtlich knapper Ressourcen darstellen und lediglich daran interessiert sind sie auszunutzen. Der Glaube daran, dass nur spezielle Beziehungen oder Glück das eigene

Fortkommen erfolgreich gestalten können, spiegelt die Wahrnehmung von einer mangelhaften Kontrolle über das subjektive Schicksal wider und wirkt sich entsprechend negativ auf die Entwicklung Generalisierten Vertrauens aus. Darüber hinaus nehmen diese Menschen die Welt als einen bedrohlichen Ort wahr, über den sie nur wenig Kontrolle besitzen. (Rosenberg (1956); Rothstein/Uslaner (2005), S. 69; Seligman (1991), S. 5; Uslaner (2001), S. 106; Uslaner (2002), S. 79, S. 95, S. 109f; Uslaner (2004), S. 502; Uslaner (2005), S. 77; Uslaner (2008c), S. 733; Uslaner (2013), S. 629ff)

Personen, die davon ausgehen, ein hohes Maß an Kontrolle über ihr eigenes Leben zu haben, sollten demnach über ein stärker ausgeprägtes Generalisiertes Vertrauen verfügen. Dagegen sollten Personen, die der Meinung sind, dass die Gestaltung des Lebensweges vornehmlich durch äußere Kräfte determiniert ist und nur geringe bis keine Einflussmöglichkeiten bestehen, ein eher niedriges Generalisiertes Vertrauen aufweisen. Es ergeben sich somit die folgenden Statements zur Erfassung dieses Aspekts Generalisierten Vertrauens:

- *Wenn man sich anstrengt, wird man in der Regel seine Lebensziele erreichen.* (Philosophies of Human Nature Original 84-item scale)
- *Der Durchschnittsmensch ist weitestgehend Herr über sein eigenes Schicksal.* (Philosophies of Human Nature Original 84-item scale)
- *Die meisten Menschen sind sich gar nicht bewusst, wie sehr ihr Leben von Zufällen beeinflusst wird.* (I-E-Scale)
- *Die meisten Menschen haben einen guten Eindruck davon, was ihre Stärken und Schwächen sind.* (Philosophies of Human Nature Original 84-item scale)

6.1.1.4.2.3 Gemeinsames Schicksal/Gemeinsame Werte/Ähnlichkeit

Generalisiertes Vertrauen spiegelt einen Sinn für soziale Solidarität wider. So gehen vertrauensvolle Menschen davon aus, dass die verschiedenen Gruppen innerhalb einer Gesellschaft im Grunde ein gemeinsames Schicksal teilen und im Wesentlichen der gleichen Wertegemeinschaft angehören. Menschen, die ein stark ausgeprägtes Generalisiertes Vertrauen aufweisen, nehmen sich als Teil einer größeren sozialen Ordnung wahr, innerhalb derer alle Menschen ein gemeinsames Los und die gleichen fundamentalen Werte miteinander teilen. Aus dieser Gemeinsamkeit ergibt sich eine Verbindung mit den anderen Gesellschaftsmitgliedern, die deren Interessen für uns relevant erscheinen lässt. Es wird dementsprechend als wenig riskant angesehen, anderen Personen und sogar Fremden mit Vertrauen zu begegnen. Personen, die ein hohes Level an Generalisiertem Vertrauen aufweisen, suchen dabei sogar aktiv nach Gemeinsamkeiten mit Personen, die aus anderen Lebenszusammenhängen stammen und die andere Weltanschauungen besitzen. Auch Hilfsbereitschaft gegenüber Fremden kommt für Personen mit einem ausgeprägten generalisierten Vertrauen, aufgrund der Annahme, Teil der gleichen Wertegemeinschaft zu sein, einem moralischen Imperativ gleich.

Personen, die ein niedriges Generalisiertes Vertrauen aufweisen, betrachten Fremde hingegen mit Argwohn und konzentrieren ihr Vertrauen eher auf Personen in ihrem unmittelbaren Umfeld. Sie vermeiden Interaktionen mit Personen, die ihnen unähnlich sind, und neigen eher dazu, sich von der Gesellschaft abzugrenzen.

(Delhey/Newton (2005); Rosenberg (1956), S. 694; Rothstein/Uslaner (2005), S. 42, S. 45, S. 52; Uslaner (2001), S. 104, S. 106; Uslaner (2002), S. 1f, S. 15, S. 26ff, S. 109f, S. 193; Uslaner (2004), S. 501; Uslaner (2005), S. 78; Uslaner (2008a), S. 103; Uslaner (2008c), S. 726; Uslaner (2012), S. 7; Uslaner (2013), S. 632)

Ein hohes Maß an Generalisiertem Vertrauen sollten somit solche Personen aufweisen, die sich eher als Teil einer weltumspannenden Wertegemeinschaft wahrnehmen und aufrichtig am Schicksal anderer Menschen interessiert sind. Es ergeben sich deshalb die folgenden Statements zur Erfassung dieses Aspekts Generalisierten Vertrauens:

- *Alle Menschen teilen allgemeine Ziele und normative Überzeugungen.* (Eigene Entwicklung)
- *Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.* (Philosophies of Human Nature Original 84-item scale)
- *Erst wenn man bekommen hat, was man vom Leben erwartet, sollte man sich mit der Ungerechtigkeit in der Welt beschäftigen.* (Mach V Attitude Inventory Version 2)
- *Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.* (Revised Philosophies of Human Nature Scale)
- *Es ist schwer einschätzbar, wie sich das Handeln der Menschen von Situation zu Situation verändert.* (Philosophies of Human Nature Original 84-item scale)

6.1.1.4.2.4 Misanthropie

Misanthropie beschreibt die Überzeugung, dass Menschen im Allgemeinen keine guten Absichten haben und eher dazu neigen andere auszunutzen. Misanthropen nehmen dementsprechend andere Menschen grundsätzlich als eher vertrauensunwürdig, unfair und weniger hilfsbereit wahr. Sie gehen davon aus, dass andere Menschen vornehmlich am eigenen Wohlergehen interessiert sind und sich nicht um andere sorgen. Insgesamt herrscht eher eine negative Weltwahrnehmung vor. Sie weisen somit eher ein geringes Vertrauen in die menschliche Natur auf, was einem eher gering ausgeprägten generalisierten Vertrauen entspricht. Dagegen nehmen Personen, die ein stark ausgeprägtes Generalisiertes Vertrauen aufweisen, die Welt als einen freundlichen Ort wahr. Andere Menschen werden als grundlegend vertrauenswürdig betrachtet und ihnen eine fundamentale Hilfsbereitschaft unterstellt. Außerdem wird davonausgegangen, dass die menschliche Natur weniger selbstsüchtig veranlagt, als altruistisch ausgelegt ist.

(Brehm/Rahn (1997), S. 1003; Smith (1997), S. 171f; Uslaner (2002), S. 26ff, S. 192; Welzel/Delhey (2015), S. 877)

Entsprechend dieser Gegensätze wurde dieser Aspekt Generalisierten Vertrauens mit Hilfe der nachfolgenden Items abgebildet:

- *Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken.* (Interpersonal Trust Scale)
- *Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.* (Mach IV Version 1)
- *Es ist möglich, in jeder Hinsicht ein guter Mensch zu sein.* (Mach IV Version 1)
- *Wenn man einen Gegenstand an eine andere Person verleiht, wird man diesen intakt zurückbekommen.* (Eigene Entwicklung)

- *Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden.* (Philosophies of Human Nature Original 84-item scale)

6.1.1.4.2.5 Vertrauensbereitschaft, Hilfsbereitschaft und Fairness

Die nachfolgenden Items orientieren sich an der Faktorstruktur der *Trust in People-Scale* und gliedern sich in deren Einzelkomponenten *Vertrauensbereitschaft*, *Hilfsbereitschaft* und *Fairness*. (Kunz (2004), S. 204f; Robinson/Shaver/Wrightsman (1991), S. 406ff)

Generalisiertes Vertrauen betrifft die grundsätzliche *Vertrauensbereitschaft* von Individuen unabhängig von situativen Spezifika. Die Ausprägung der Vertrauensbereitschaft spiegelt dabei auch die Wahrnehmung der Vertrauenswürdigkeit der Mitmenschen wider. Diese generalisierte Vertrauensbereitschaft kann dabei im Sinne einer *culture of trust* durch eine grundlegende Tendenz zu wechselseitigem Vertrauen innerhalb einer Gesellschaft determiniert sein. (Kunz (2004), S. 204; Lahno (2002), S. 210; Sztompka (1999), S. 119ff) Die Komponente Vertrauensbereitschaft wird dementsprechend anhand von alltäglichen Situationen abgebildet, die Vertrauen erforderlich machen:

- *Es ist sicherer, Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.* (Trust Inventory)
- *Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen.* (Philosophies of Human Nature Original 84-item scale)
- *Es spielt keine Rolle, ob eine Wohngegend als sicher gilt, man sollte seine Wohnungstür immer abschließen, wenn man das Haus verlässt.* (Abgeleitet aus SOEP2013)
- *Wenn man dem Normalbürger eine Aufgabe gibt, so wird er sie erfolgreich erfüllen, ohne dass er überwacht werden muss.* (Philosophies of Human Nature Original 84-item scale)
- *Wenn man möchte, dass die Leute einen Auftrag korrekt ausführen, sollte man ihnen die Aufgabe detailliert erklären und sie genau beaufsichtigen.* (Philosophies of Human Nature Original 84-item scale)

Zwar betrachtet USLANER *Hilfsbereitschaft* als ein von Vertrauen zu differenzierendes Konzept, das auch losgelöst von der Bereitschaft Vertrauen zu schenken, existieren kann. (Uslaner (2002), S. 69ff) Allerdings steht *Hilfsbereitschaft* insofern mit (Generalisiertem) Vertrauen in Verbindung, als dass sie dazu in der Lage ist, eine Vertrauensbeziehung zu initiieren. So stellt die Erfahrung von Hilfsbereitschaft eine besonders positive Kooperationserfahrung dar, die als *leap of faith* bzw. einseitiger Vertrauensvorschuss fungieren kann und möglicherweise eine Erweiterung des individuellen Vertrauens über Gruppengrenzen hinweg bewirkt. Dies kann insbesondere dann der Fall sein, wenn man wiederholt Empfänger von Hilfsbereitschaft wird. (Newton 1999a), S. 171; Stolle (1998), S. 503f) Entsprechend wurde die Komponente Hilfsbereitschaft anhand von Items operationalisiert, die altruistisches Verhalten thematisieren:

- *Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist.* (Revised Philosophies of Human Nature Scale)
- *Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.* (Philosophies of Human Nature Original 84-item scale)

Auch bezüglich *Fairness* vertritt USLANER die Auffassung, dass es sich um eine von Vertrauen zu unterscheidendes Konstrukt handelt. Die *Fairness* anderer Personen lässt sich hiernach nur auf der Basis detaillierten Wissens über deren Wertekatalog beurteilen. (Uslaner (2002), S. 70) KUNZ geht dagegen davon aus, dass Generalisiertes Vertrauen u.a. eine generalisierte Erwartung hinsichtlich der *Fairness* generalisierter Anderer beinhaltet. (Kunz (2004), S. 204) Es wird angenommen, dass Personen, denen Vertrauen entgegengebracht wird, einer Reziprozitätsnorm folgen und erwiesene Gefallen spiegeln werden. Darüber hinaus spielt Fairness im Sinne einer institutionellen Fairness eine große Rolle für die Verbreitung generalisierter Vertrauensdispositionen innerhalb einer Gesellschaft. (Nannestad (2008), S. 424; Park (2012), S. 5) Die Wahrnehmung von *Fairness* wird anhand von Statements gemessen, die Aufrichtigkeit und Reziprozität zum Gegenstand haben:

- *Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.* (Philosophies of Human Nature Original 84-item scale)
- *Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.* (Inventar zur Erfassung des interpersonellen Vertrauens (IIV))
- *Man kann ein Versprechen ruhig brechen, wenn es für einen selbst vorteilhaft ist.* (Machiavellismus-Skala)
- *Man sollte nur dann den wahren Grund seiner Absichten zeigen, wenn es einem nützt.* (Mach V Attitude Inventory Version 2)
- *Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.* (Interpersonal Trust Scale)

6.1.1.4.2.6 Formelle/Informelle Netzwerke

Formelle und informelle soziale Netzwerke ermöglichen Interaktion, Kommunikation und Kooperation von Individuen. Durch die wiederholte *face-to-face*-Interaktion innerhalb solcher Vereinigungen erlangen die Mitglieder wechselseitig Informationen übereinander und werden auf diese Weise in die Lage versetzt, dem jeweils anderen Vertrauenswürdigkeit zu zuschreiben. Die Erlangung einer vertrauenswürdigen Reputation erhöht in der Folge die Chance für Kooperationen mit anderen Netzwerkmitgliedern. Den Mitgliedern einer solchen Vereinigung werden auf diese Weise die *civic virtues of trust* vermittelt und sie erlernen Heuristiken zur Vertrauensvergabe. Gleichzeitig entwickelt sich ein Sinn für Gemeinschaftsinteressen und Gemeinschaftsgüter.

Die Übertragung dieser Inhalte auf die gesamte Gesellschaft und die Manifestation in Form von Generalisiertem Vertrauen erfolgt im Wesentlichen durch die Übertragung positiver Kooperationserfahrungen, die im Rahmen des Netzwerks gesammelt wurden, auf den gesamtgesellschaftlichen Kontext. Dabei können verschiedene Mechanismen zum Einsatz kommen. Die Mitgliedschaft in einem Netzwerk erhöht die Kooperationsbereitschaft, weshalb auch außerhalb des Netzwerkes öfter Vertrauensbeziehungen eingegangen werden. Auf diese Weise sorgt die Interaktion mit Nicht-Mitgliedern für eine Vermittlung der *civic virtues of trust* in die Gesellschaft als Ganzes. Hierbei kann auch die Projizierung der eigenen Vertrauenswürdigkeit auf Fremde bzw. die Anwendung erfahrungsbasierten Vertrauens auf unbekannte Personen eine Rolle spielen. Ebenso ermöglichen aber auch überlappende und horizontal organisierte Netzwerke, die eine heterogene Gruppenzusammensetzung aufweisen,

eine Verbreitung Generalisierten Vertrauens. Netzwerkmitglieder können hier mit Personen aus verschiedenen gesellschaftlichen Kontexten in Kontakt treten und so Erfahrungen mit gruppenübergreifender Reziprozität sammeln. Auch die inhaltliche Ausrichtung der Vereinigung kann hierbei einen positiven Einfluss haben. Ebenso spielen Freundschaften eine große Rolle bei der Entwicklung und Aufrechterhaltung Generalisierten Vertrauens. Regelmäßiger Kontakt mit Freunden und Bekannten beeinflusst das Ausmaß Generalisierten Vertrauens nachweislich positiv. (Delhey/Newton (2002), S. 6ff; Delhey/Newton (2004), S. 163; Newton (1999a), S. 171ff; Newton (1999b), S. 11, S. 15; Newton (2001), S. 206f; Newton/Norris (2000), S. 66; Putnam (1995a), S. 67, S. 73; Putnam (1995b), S. 20, S. 29; Putnam (2000), S. 19, S. 136f; Putnam/Leonardi/Nannetti (1993), S. 169, S. 173f; Stolle (1998), S. 500ff; Stolle (2001a), S. 204f; Stolle (2001b), S. 119ff, S. 130; Stolle (2003), S. 24ff; Stolle/Hooghe (2003), S. 28; Stolle/Rochon (1998), S. 57ff; Stolle/Rochon (1999), S. 192f; Stolle/Rochon (2001), S. 144, S. 152ff; Wollebæk/Selle (2002), S. 37f; Wollebæk/Selle (2007), S. 4)

Die Items zu diesem Aspekt Generalisierten Vertrauens gliedern sich in die Bezugspunkte formelle und informelle Netzwerke. Ziel war es, die Erfahrungen mit Netzwerken dieser Art abzubilden und hieraus Rückschlüsse auf das Ausmaß Generalisierten Vertrauens bei den Befragten zu ziehen. Entsprechend sollten positive Erfahrungen ein stark ausgeprägtes Generalisiertes Vertrauen anzeigen. Eher auf formelle Netzwerke bezogen waren dabei diese Items:

- *Der persönliche Einsatz für Vereine und Organisationen stellt eine erfüllende Freizeitbeschäftigung dar.* (Eigene Entwicklung)
- *Das Engagement in politischen Parteien sollte innerhalb der Gesellschaft zu einem allgemeingültigen Wert werden.* (Abgeleitet aus Sozio-politische Einstellungen (Politisches Engagement))

Dagegen waren die folgenden Items eher auf informelle Netzwerke bezogen:

- *Es ist nicht schwer, Menschen zu finden, die einem beim Umzug helfen würden.* (Eigene Entwicklung)
- *Heutzutage ist es nicht schwer Freunde zu finden.* (Eigene Entwicklung)

6.1.1.4.2.7 Institutionen/Soziale Ordnung

Der Einfluss politischer Institutionen auf Generalisiertes Vertrauen ergibt sich aus ihrer ordnungsstiftenden Kraft. So sorgen politische Institutionen und hierbei insbesondere sogenannte Implementationsinstitutionen wie Judikative und Administration durch die Überwachung und Gewährleistung der Rechtsordnung bzw. institutioneller Regeln für die Aufrechterhaltung sozialer Ordnung. Sie verpflichten damit sämtliche Gesellschaftsmitglieder innerhalb einer bestimmten normativen Rahmung zu handeln, wobei jeder gleiche Rechte und Pflichten besitzt. Durch die Einbettung des Handelns jedes Einzelnen in einen institutionellen Rahmen wird dann das Verhalten wechselseitig antizipierbar. Auf diese Weise kann auch Fremden Vertrauen entgegengebracht werden, solange sich diese im gleichen institutionellen Kontext bewegen wie man selbst.

Bewerkstelligt wird dies durch die Wahrnehmung von Institutionen als unparteiisch, fair, neutral, objektiv, unbestechlich, universalistisch, effizient und deviantes-Verhaltens-sanktionierend. So werden vom Handeln der Institutionen Rückschlüsse auf das Handeln der Mitbürger gezogen, wobei die von den Institutionen propagierten Werte und Normen als Grundlage für die Bewertung des Verhaltens der Mitmenschen dienen. Gleichzeitig kann angenommen werden, dass Verstöße gegen die geltende Ordnung geahndet werden, was die Wahrnehmung persönlicher Sicherheit steigert. Die Befolgung der obengenannten Prinzipien durch die Institutionen bzw. das Vertrauen darauf, dass die Prinzipien befolgt werden, hat somit unmittelbaren Einfluss auf die Verbreitung Generalisierten Vertrauens innerhalb einer Gesellschaft.

Implementationsinstitutionen sind hierbei deshalb so wirksam, weil sie durch Ihre Beständigkeit über einzelne Legislaturperioden hinaus Einfluss auf die Sozialisation der Bürger nehmen können. Darüber hinaus besteht durch die Möglichkeit unmittelbarer Interaktion mit Vertretern dieser Institutionen ein direkter Kontakt mit dem Bürger, was die Wirksamkeit als Vorbilder für gemeinschaftsorientiertes Handeln verstärkt.

(Delhey/Newton (2005), S. 313; Freitag/Bühlmann (2005), S. 577, S. 581ff; Freitag/Bühlmann (2009), S. 1539, S. 1543ff; Levi (1998), S. 86ff; Newton (2008), S. 255f; Newton/Norris (2000), S. 53; Newton/Zmerli (2011), S. 169; Offe (2000), S. 127f, S. 130; Offe (2001a), S. 265f, S. 274, S. 279ff; Stolle/Rothstein (2001), S. 7ff, S. 16f, S. 38; Stolle/Rothstein (2002), S. 11ff, S. 29; Stolle/Rothstein (2003), S. 194f, S. 199f; Stolle/Rothstein (2007), S. 11ff, S. 16, S. 29; Stolle/Rothstein (2008a), S. 283f; Stolle/Rothstein (2008b), S. 446; Uslaner (2002), S. 43ff, S. 80f, S. 148)

Personen, die ein stark ausgeprägtes Generalisiertes Vertrauen aufweisen, sollten somit davon ausgehen, dass staatliche Institutionen alle Menschen gleich und fair behandeln und dass die gesellschaftlichen Regeln für alle Gesellschaftsmitglieder gleichermaßen Gültigkeit besitzen. Es wurden daher die folgenden Items verwendet, um diesem Aspekt Generalisierten Vertrauens gerecht zu werden. Dabei zeigt die Zustimmung zu den positiv formulierten Items ein hohes Maß an Generalisiertem Vertrauen an:

- *Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden.* (Interpersonal Trust Scale)
- *Hat man in irgendeiner Angelegenheit mit staatlichen Behörden zu tun, so wird man in der Regel fair und gerecht behandelt.* (Abgeleitet aus Behandlung durch Behörden)

Während bezüglich der negativ formulierten Items eine Ablehnung ein höheres Maß an Generalisiertem Vertrauen anzeigt:

- *Wir verlieren langsam unsere persönliche Freiheit, weil die Behörden sich immer weiter ausdehnen.*
- *Der Mensch sollte nur diejenigen staatlichen Beschränkungen seiner Freiheit befolgen, die er selber als vernünftig ansieht.* (Machiavellismus-Konservatismus (MK-Skala))

6.1.1.4.2.8 Heterogenität/Diversität

Eine heterogene Zusammensetzung des individuellen sozialen Umfeldes kann sowohl negative als auch positive Effekte auf die Ausgestaltung Generalisierten Vertrauens haben. Während Konflikt- und Konstrukttheorie postulieren, dass soziale Nähe zu Angehörigen anderer gesellschaftlicher Gruppen (z.B: Religion, Nationalität, Ethnie) feindseliges Verhalten gegenüber der *outgroup* zur Folge hat und letztlich eine Reduktion des Generalisierten Vertrauens bewirkt, nimmt die Kontakttheorie an, dass die Interaktion und Kommunikation mit einer oder mehreren *outgroups* dazu beiträgt wechselseitige Ressentiments abzubauen und Toleranz zu fördern, was in letzter Instanz eine Stärkung des Generalisierten Vertrauens nach sich zieht. Handlungsleitend für die Operationalisierung des Faktors „Heterogenität/Diversität“ war die Kontakttheorie, sodass eine positive Einstellung gegenüber einer heterogenen Gesellschaftsstruktur ein höheres Generalisiertes Vertrauen anzeigt.

(Gundelach (2014a), S. 55ff, S. 66; Gundelach (2014b), S. 128; Gundelach/Freitag (2014), S. 1240; Gundelach/Trautmüller (2014), S. 599ff); Rothstein/Uslaner (2005); Uslaner (2013), S. 632f)

Der Faktor enthält dabei sowohl negativ als auch positiv formulierte Items. Bei dem positiv formulierten Item zeigt die Zustimmung der Probanden ein höheres Generalisiertes Vertrauen an:

- *Im Allgemeinen handeln alle Menschen auf der Welt nach den gleichen Werten und Normen.* (Eigene Entwicklung)

Dagegen zeigt bei den negativ formulierten Items eine ablehnde Haltung gegenüber dem jeweiligen Statement ein höheres Maß Generalisierten Vertrauens an:

- *Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinandertreffen, ist von andauernden Konflikten geprägt.* (Eigene Entwicklung)
- *Der Umgang mit Personen aus dem näheren persönlichen Umfeld gestaltet sich einfacher, als der Umgang mit Fremden.* (Eigene Entwicklung)
- *Eine Kooperation von Personen, die sich ziemlich ähnlich sind, wird erfolgreicher sein, als eine Kooperation von Personen, die sich stark voneinander unterscheiden.* (Eigene Entwicklung)

6.1.1.4.3 Kognitiver Pretest

Ein Kerncharakteristikum indirekter Messverfahren ist die geringe Aufmerksamkeit der Probanden für das jeweilige Zielkonstrukt und die daraus resultierende Verhinderung vorsätzlicher Modifikationen des Antwortverhaltens. Im vorliegenden Fall sollte dies durch eine Verschleierung der Messintention des Fragebogens realisiert werden. Um zu prüfen, ob die verwendeten Items tatsächlich eine geringe Augenscheinvalidität hinsichtlich der Zieldimension „Generalisiertes Vertrauen“ aufwiesen, wurde die Methode des lauten Denkens (*Think-Aloud-Technik*), die im Rahmen qualitativer kognitiver Pretests zum Einsatz kommt, für den Einsatz innerhalb eines standardisierten Fragebogens adaptiert. Die *Think-Aloud-Technik* beinhaltet eine Aufforderung zum „lauten Denken“ bzw. zur Verbalisierung sämtlicher

Gedankengänge, die zur Antwort führen (*Concurrent Think Aloud*) bzw. geführt haben (*Retrospective Think Aloud*). Das hier gewählte Vorgehen entspricht dabei weitgehend dem *Retrospective Think Aloud*, bei dem die Probanden im Anschluss an die Beantwortung einer Frage gefragt werden, aufgrund welcher Überlegungen ihre Antwort zustande kam. (Häder (2015), S. 402; Lenzner/Neuert/Otto (2014), S. 3; Prüfer/Rexroth (2000), S. 8f)

So wurde im Anschluss an die Fragebatterie zur Erfassung Generalisierten Vertrauens zunächst Folgendes gefragt:

„Was glauben Sie, wurde mit Hilfe der vorangegangenen Statements abgefragt bzw. welches Konstrukt sollte Ihrer Meinung mit den vorangegangenen Aussagen gemessen werden? Bitte fassen Sie Ihre Annahme nach Möglichkeit in einem oder mehreren Schlagworten zusammen.“

Wurde hierbei ein von (Generalisiertem) Vertrauen abweichendes Konstrukt genannt, wurde angenommen, dass die verwendeten Items eine ausreichend geringe Augenscheinvalidität aufwiesen. Wurde von den Probanden dagegen (Generalisiertes) Vertrauen als Zieldimension identifiziert, so deutete dies daraufhin, dass die Verschleierung der Messintention nicht erfolgreich war.

Um zu ermitteln, welche Items im Speziellen für die Aufdeckung der Messintention verantwortlich waren, wurde hiernach außerdem die nachfolgende Frage gestellt:

„Was hat Sie zu Ihrer Antwort auf die vorangegangene Frage bewogen bzw. woran glauben Sie die Intention des Fragebogens erkannt zu haben? Bitte geben Sie nach Möglichkeit das jeweilige Statement sinngemäß oder wörtlich wieder.“

Die hier genannten Items standen unter besonderer Beobachtung und sollten im Zuge der Itemanalyse bevorzugt eliminiert werden.

6.1.1.4.4 Radiusproblem

Das Radiusproblem stellt die größte Problematik innerhalb der Vertrauensmessung dar und befasst sich mit dem Personenkreis, der als Referenz für das individuelle Vertrauenslevel herangezogen wird. Um zu prüfen, ob die Probanden bei der Bearbeitung der Skala zur Erfassung Generalisierten Vertrauens eher der *ingroup*, d.h. Personen, die ihnen unmittelbar bekannt sind, oder ihrem direkten sozialen Umfeld entstammen, oder die *outgroup*, d.h. Personen, über die nur wenige Informationen bekannt sind, bzw. Fremde als Referenz für ihren Vertrauenslevel nutzen, wurde im Anschluss an den kognitiven Pretest folgende Frage gestellt:

„An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht? Bitte wählen Sie nur eine der Antwortalternativen aus.“

Die Antwortoptionen *„Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)“*, und *„Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)“*, entsprachen dabei der *ingroup* als Bezugsgruppe. Die Auswahl einer dieser Antwortalternativen

6. Pilotstudie

würde daher anzeigen, dass mit Hilfe der Vertrauensskala eher Partikulares Vertrauen abgebildet wird.

Die Antwortoptionen „Personen, die Sie erst ein einziges Mal getroffen haben“, „Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)“, und „Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)“ zeigten dagegen die *outgroup* als Bezugsgruppe an. Die Auswahl einer dieser Antwortalternativen würde daher anzeigen, dass generalisierte Andere als Referenz herangezogen wurden und die Vertrauensskala vornehmlich Generalisiertes Vertrauen abbildet.

6.1.1.4.5 Soziale Erwünschtheit

Unter sozialer Erwünschtheit versteht man in der Regel ein Antwortverhalten seitens der Befragten, das sich an Normen und Erwartungen für positiv sanktioniertes Verhalten orientiert. Solches Antwortverhalten tritt insbesondere in Situationen auf, in denen man sich besonders gut darstellen möchte, in denen man negative Konsequenzen aufgrund der eigenen Antwort befürchtet oder in denen man ein positives Bild von sich selbst bewahren möchte. Außerdem spielt eine Rolle, ob die Thematik der Befragung als besonders heikel, kontrovers oder stigmatisierend wahrgenommen wird. Sozialerwünschtes Antwortverhalten (*social desirability*) kann somit definiert werden als der Versuch positives Verhalten oder besonders vorteilhafte Eigenschaften und Merkmale der eigenen Person hervorzuheben, während gleichzeitig negativ assoziiertes Verhalten oder ungünstige Eigenschaften und Merkmale verschleiert werden. Hieraus können sich mitunter erhebliche Verzerrungen in den Daten empirischer Untersuchungen ergeben, sodass soziale Erwünschtheit die Validität der Ergebnisse bedrohen kann. (Diekmann (2011), S. 443; Bortz/Döring (2006), S. 232ff; Bühner (2006), S. 60ff; Schnell/Hill/Esser (2008), S. 355f)

Erfasst wurde sozial erwünschtes Antwortverhalten in dieser Studie mit Hilfe einer vier Items umfassenden Kontrollskala, die auf einer von MARLOWE/CROWNE entwickelten SD-Skala basiert und von TIMAEUS/LÜCK ins Deutsche übersetzt wurde. (Lück/Timaeus (1969); Lück/Timaeus (2014))

Die Kurzskala umfasst insgesamt vier Items, von denen zwei positiv und zwei negativ formuliert waren. Als Antwortalternativen wurden „stimmt“ und „stimmt nicht“ angeboten⁴⁰.

Tab. 6.3: Soziale Erwünschtheit SDS-CM (Kurzskala)

Item	Antwortschablone
Ich sage immer, was ich denke.	R
Ich bin manchmal ärgerlich, wenn ich meinen Willen nicht bekomme.	F
Ich bin immer gewillt, einen Fehler, den ich mache, auch zuzugeben.	R
Ich habe gelegentlich mit Absicht etwas gesagt, was die Gefühle des anderen verletzen könnte.	F

Auswertungsschlüssel:

F = bei Ankreuzen der Antwortalternative „stimmt nicht“ erhält die Person einen Punkt für sozial erwünschtes Antwortverhalten.

R = bei Ankreuzen der Antwortalternative „stimmt“ erhält die Person einen Punkt für sozial erwünschtes Antwortverhalten.

Jede Antwort in Richtung der Zieldimension „Soziale Erwünschtheit“ wird dabei mit einem Punkt bewertet und die Punkte über die gesamte Skala aufsummiert. Bei vollständig sozial

⁴⁰ Im Original wird eine dichotome Antwortskala mit den Kategorien „richtig“ bzw. „falsch“ verwendet.

erwünschtem Antworten ist somit eine maximale Punktzahl von 4 Punkten pro Befragtem für die Kurzskala erreichbar.

6.1.1.4.6 Sozioökonomischer Status

Die Items v5.6 bis v5.8 betreffen den sozioökonomischen Status der Befragten. Aus empirischen Studien ergab sich, dass ein hohes Einkommen bzw. materielle Sicherheit positiv mit Generalisiertem Vertrauen korrelieren. Menschen, die keinen materiellen Einschränkungen unterliegen, sind eher bereit, sich den Risiken einer Vertrauensbeziehung auszusetzen, da sie mögliche Verluste, die sich aus einem Vertrauensmissbrauch ergeben können, leichter zu kompensieren vermögen. Außerdem zeigt sich, dass finanziell gut gestellte Personen eine optimistischere Weltwahrnehmung haben und auch ihre Mitmenschen in einem eher positiven Licht sehen, was ebenfalls eine Steigerung des individuellen Ausmaßes Generalisierten Vertrauens bewirkt.

Gleichzeitig kann ein niedriges Einkommen zu der Überzeugung führen, dass man sich weniger als Teil der Gesellschaft empfindet, was in einem individuell niedrigen Generalisierten Vertrauen resultieren kann.

Betrachtet man die gesamtgesellschaftliche Ebene, so stellt soziale Polarisierung in Form von Einkommensungleichheit eine der Hauptursachen für ein gering ausgeprägtes Generalisiertes Vertrauen innerhalb eines Gesellschaftsgefüges dar. (Alesina/LaFerrara (2000), S. 8; Alesina/LaFerrara (2002), S. 219; Bjørnskov (2006), S. 7; Bjørnskov (2008), S. 276; Delhey/Newton (2004), S. 163; Delhey/Newton (2005), S. 312; Delhey/Newton/Welzel (2011), S. 790; Smith (1997), S. 180ff, S. 189; Uslaner (2002), S. 38, S. 85, S. 90; Whiteley (1999), S. 40)

Da es sich bei den Befragten um Studierende handelt und diese in den meisten Fällen noch nicht über ein eigenes Einkommen verfügen, wurde neben den monatlich zur Verfügung stehenden Geldmitteln auch die Berufe der Eltern abgefragt. Letzteres soll Rückschlüsse auf die soziale Herkunft der Befragten ermöglichen und die materielle Sicherheit bzw. die Verfügbarkeit materieller Ressourcen für die Befragten dokumentieren. Eine Klassifizierung des sozioökonomischen Status auf Basis der Berufe der Eltern ist anhand von entsprechenden Skalen realisierbar. (Wolf (1995), S. 104ff)

6.1.1.4.7 Mobilität und soziale Gemeinschaft

Die Items v5.9 bis v5.11 betreffen die geographische Mobilität und die damit verbundene Stabilität des sozialen Umfeldes. Mobilität wirkt sich in doppelter Hinsicht auf Generalisiertes Vertrauen aus. So kann sich ein Mangel an Vertrauen auf der einen Seite daraus ergeben, dass man noch nicht lange am derzeitigen Wohnort lebt bzw. bereits mehrfach umgezogen ist und man die dort lebenden Menschen noch nicht gut genug kennt, um ihnen zu vertrauen. Auf der anderen Seite kann die extreme Mobilität des eigenen sozialen Umfeldes dazu führen, dass man kaum die Möglichkeit hat, eine Vertrauensbeziehung aufzubauen. In einem solchen Fall sorgt die Flüchtigkeit der sozialen Gemeinschaft dafür, dass nur wenige Gelegenheiten zur sozialen Interaktion gegeben sind und sich auf diese Weise ein Mangel an Vertrauen einstellt. Aus empirischen Untersuchungen wurde ersichtlich, dass sich insbesondere die Stabilität des sozialen Umfeldes positiv auf die individuelle Vertrauensneigung auswirkt. (Alesina/LaFerrara (2000), S. 11ff; Alesina/LaFerrara (2002), S. 223ff)

6.1.1.4.8 Anmerkungen

Den Abschluss des Fragebogens bildete eine offene Frage, die den Probanden ermöglichte, Anmerkungen und Verbesserungsvorschläge zu artikulieren.

6.1.1.5 Validitätsbedrohungen und systematische Verzerrungen

Im Rahmen von Befragungen können eine Reihe von Störfaktoren Verzerrungen der jeweiligen Messergebnisse bewirken. Nachfolgend sollen potentielle Störfaktoren, die bisher noch nicht diskutiert wurden, aber trotzdem besondere Relevanz für die Pilotstudie besitzen, beschrieben werden.

6.1.1.5.1 Reihenfolgeeffekte

Reihenfolgeeffekte beschreiben Antwortverzerrungen, die sich aus der Abfolge der Items bzw. deren relativer Position zueinander innerhalb eines Fragebogens ergeben. Im Kontext der Pilotstudie besteht insbesondere die Möglichkeit des Auftretens von sogenannten *Primacy-Recency-Effekten* und *Rapport-Fatigue-Effekten*. (Scholl (2015), S. 217f; Sudman/Bradburn/Schwarz (1996), S. 130) So weisen die Probanden möglicherweise mit den Statements am Anfang der Itematterie zur Erfassung Generalisierten Vertrauens ein größeres Einverständnis auf als mit den Statements an deren Ende oder umgekehrt. Gleichzeitig könnte es auch passieren, dass die Probanden die Items am Ende der Fragebatterie weniger reflektiert beantworten als die Items am Anfang oder umgekehrt. In beiden Fällen werden die Bias durch die erhebliche Länge der fraglichen Itematterie moderiert.

6.1.1.5.2 Formale Antwortstile und Antworttendenzen

Antwortstile bzw. Antworttendenzen beschreiben ein Antwortverhalten, das unabhängig vom Inhalt des jeweiligen Items ist und vornehmlich in der Persönlichkeit der Probanden oder den situativen Begebenheiten begründet ist. Im Rahmen der Pilotstudie können dabei insbesondere eine Ja- oder Nein-Sage-Tendenz, eine Tendenz zur Mitte, eine Tendenz zu extremen Urteilen, der Milde-Härte-Fehler sowie *non-attitudes* oder Pseudomeinungen auftreten. (Möhring/Schlütz (2010), S. 57ff; Schnell/Hill/Esler (2008), S. 353f; Scholl (2015), S. 225ff) So ist anzunehmen, dass die Probanden aufgrund der Länge der Fragebatterie zur Erfassung Generalisierten Vertrauens zur Verwendung entsprechender Antwortmuster neigen.

6.1.1.6 Auswertungskonzept

Die Analyseschritte, anhand derer die Selektion der Items vorgenommen wurde, die im Rahmen der Pilotstudie zu einer Objektiv-Skala zusammengefasst werden sollten, orientieren sich im Wesentlichen an den Empfehlungen und Vorgaben von BÜHNER, FISSENI, LIENERT/RAATZ und MOOSBRUGGER/KELAVA für Item-, Reliabilitäts- und Validitätsanalysen sowie den sogenannten *Jackson Principles* (Butler (1991), S. 649ff). Letztere definieren die Abfolge der Maßnahmen, die

zur Konstruktion und Validierung einer Skala notwendig sind und gliedern sich in die nachfolgenden Phasen: a) theoriebasierte Definition des Zielkonstruktes, b) Bestimmung von Reliabilität und Homogenität der Skala, c) Kontrolle reaktiver Effekte und d) Bestimmung konvergenter und diskriminanter Validität. Die Itemanalyse wird dabei zwischen den Schritten a) und b) eingefügt und beinhaltet eine erste Selektion geeigneter Items anhand von Itemschwierigkeit und Itemtrennschärfe.

Die Definition des Zielkonstruktes „Generalisiertes Vertrauen“ und die Ableitung entsprechender Messvorschriften erfolgte bereits im Kontext der theoretischen Aufarbeitung sowie der Operationalisierung und ist gleichzusetzen mit der Bestimmung der Inhaltsvalidität der Items. (Butler (1991), S. 649f)

Im Rahmen der Itemanalyse erfolgte eine erste Selektion bzw. Elimination von Items infolge der Bestimmung von Itemschwierigkeit und Itemtrennschärfe. Ziel war die Erstellung eines möglichst homogenen Tests bzw. einer Skala, die nur eine einzelne Dimension abbildet. Zu diesem Zweck sollten nur solche Items beibehalten werden, deren Schwierigkeitsindex im Intervall $P_m = .40-.60$ liegt und deren Itemtrennschärfe gleichzeitig möglichst maximal wird bzw. mindestens im Intervall $r_{i(x-i)} = .40-.70$ liegt. Vorrangiges Selektionskriterium ist dabei allerdings die Itemtrennschärfe, da Items mit hoher Trennschärfe eine besonders gute Passung mit der jeweiligen Zieldimension aufweisen. Es sollten deshalb auch solche Items zugelassen werden, die vergleichsweise extreme Itemschwierigkeiten aufweisen bei gleichzeitig hohen Trennschärfen. Dies hat den Nebeneffekt, dass die ermittelte Skala eine höhere Differenzierungsvielfalt erhält und besser zwischen Personen mit außergewöhnlichen Merkmalsausprägungen trennen kann. (Bortz/Döring (2016), S. 475ff; Fisseni (1997), S. 46f, S. 60ff; Lienert/Raatz (1998), S. 115; Moosbrugger/Kelava (2012), S. 86)

Die im Zuge der Itemanalyse selektierten Items werden im nächsten Schritt einer Reliabilitätsanalyse unterzogen, in deren Rahmen eine zweite Selektion bzw. Elimination stattfindet. Mit Hilfe des Verfahrens der Alpha-Maximierung wird nun die Reliabilität der Skala gesteigert, indem Items mit niedrigen Trennschärfen ($r_{i(x-i)} < .30$) so lange schrittweise ausgeschlossen werden, bis keine weitere Erhöhung des Reliabilitätskennwertes *Cronbach- α* mehr zu verzeichnen ist. Eine rudimentäre Homogenitätsprüfung der resultierenden Objektiv-Skala erfolgt anhand einer explorativen Hauptachsenanalyse. (Bortz/Döring (2016), S. 479ff; Bühner (2006), S. 138; Bühner (2011), S. 168, S. 249f)

Konvergente und diskriminante Validität lassen sich nach BÜHNER unter dem Oberbegriff Konstruktvalidität zusammenfassen. (Bühner (2011), S. 63f) Die Ermittlung dieser beiden Validitätsformen erfolgt dabei anhand der Korrelation der Objektiv-Skala mit Messinstrumenten, die die gleiche Dimension erfassen bzw. eine andere Dimension erfassen. Mangels einer parallelen Messung Generalisierten Vertrauens anhand einer alternativen Vertrauensskala, wurde die Bestimmung der Konstruktvalidität bzw. im Speziellen der konvergenten Validität in die Hauptstudie ausgelagert. Stattdessen wird allerdings die Kriteriumsvalidität, d.h. der Zusammenhang der Objektiv-Skala mit externen Kriterien, die in Verbindung mit dem theoretischen Konstrukt „Generalisiertes Vertrauen“ stehen bestimmt.

6.2 Datenauswertung

6.2.1 Itemanalyse (Objektiv-Skala)

Die Itemanalyse soll dazu dienen, die Anzahl der Items zu reduzieren und auf diese Weise eine kompakte Vertrauensskala nach den Kriterien Objektiver Tests zu erhalten. Da in der Ursprungsskala sowohl positiv wie negativ formulierte Items enthalten waren, wurden zunächst die negativ gepolten Items entsprechend umcodiert, sodass nun alle in die weiteren Analysen eingehenden Items in Schlüsselrichtung der Skala weisen. (Bühner (2011), S. 217)

Hohe Skalenwerte bedeuten somit für alle Items eine Tendenz der Befragten in Richtung „Generalisiertes Vertrauen“, während niedrige Skalenwerte eine Tendenz der Befragten invers zur Richtung dieser Zieldimension anzeigen.

Die statistische Itemanalyse erfolgt auf Basis der Itemrohwerte und betrachtet im Wesentlichen die Itemschwierigkeit, die Itemstreuung und die Itemtrennschärfe als Selektionskriterien. Auf diese Weise werden diejenigen Items aus der Rohskala eliminiert, die nicht zur Unterscheidung der Merkmalsausprägungen der Probanden beitragen bzw. das Zielkonstrukt nicht erfassen. Im Anschluss daran wird die Messgenauigkeit der resultierenden Skala anhand einer Reliabilitätsanalyse überprüft. Außerdem erfolgt eine rudimentäre Prüfung auf Eindimensionalität, um sicherzustellen, dass es sich bei dem resultierenden Messinstrument um eine homogene Skala handelt. (Bühner (2006), S. 52)

6.2.1.1 Itemschwierigkeit und Itemtrennschärfe (Objektiv-Skala)

Die sogenannte psychometrische Itemschwierigkeit beschreibt das Ausmaß der Zustimmung zu einem Item in Messrichtung der zugrundeliegenden Skala. Für Ratingskalen bzw. intervallskalierte Items wird hier in der Regel der jeweilige Itemmittelwert als Schwierigkeitsindex herangezogen. Es gilt dann:

„Ist ein Item psychometrisch schwer, werden diesem Item viele Personen der Stichprobe zustimmen, und daher wird der Mittelwert des Items hoch ausfallen. Niedrige Mittelwerte kennzeichnen psychometrisch leichte Items, da ihnen wenige Personen zustimmen werden.“ (Bühner (2011), S. 220)

Bei psychometrisch schweren Items bzw. Items mit einem hohen Schwierigkeitsindex fällt es den Befragten somit besonders leicht, in Richtung der Zieldimension „Generalisiertes Vertrauen“ zu antworten, während es den Befragten bei psychometrisch leichten Items bzw. Items mit einem niedrigen Schwierigkeitsindex besonders schwer fällt, in Richtung der Zieldimension „Generalisiertes Vertrauen“ zu antworten.⁴¹ Items, die weder zu „leicht“ noch zu „schwer“ sind bzw. eine mittlere Itemschwierigkeit aufweisen, sollten demnach am besten

⁴¹ Oft wird auch die gegenteilige Formulierung verwendet: „Ein Item ist dann „schwierig“, wenn ihm von den meisten Personen nicht zugestimmt werden kann, sodass es einen niedrigen Mittelwert aufweist. Extrem „leichte“ Items haben sehr hohe Mittelwerte.“ (Mummendey/Grau (2014), S. 123) SCHNELL/HILL/ESSER weisen allerdings daraufhin, dass es sich hierbei um die umgangssprachliche Bedeutung der Itemschwierigkeit handelt, die Irritationen vorbeugen soll. Tatsächlich ist „die „Schwierigkeit“ eines Items als der Anteil der richtig gelösten Items definiert“ (Schnell/Hill/Esser (2008), S. 182), sodass Items mit hoher Schwierigkeit leicht zu lösen sind und Items mit niedriger Schwierigkeit schwer zu lösen sind.

Dieser Umstand ist auch bei der Verwendung und Interpretation der Bewertung von *Schwierigkeitsindizes von FISSENI* zu beachten. (Bühner (2011), S. 76ff; Fisseni (1997), S. 124; Lienert/Raatz (1998), S. 73)

geeignet sein, Merkmalsdifferenzen zwischen den Befragten zu erfassen bzw. zwischen Befragten, die ein Item lösen oder nicht lösen, zu trennen. (Bühner (2006), S. 98; Moosbrugger/Kelava (2012), S. 76, S. 81, S. 83)

Die Itemtrennschärfe gibt an, wie gut ein Item die jeweilige Zieldimension misst. Für Intervallskalen wird als Trennschärfe der Pearson-Produkt-Moment-Korrelationskoeffizient verwendet. Die Trennschärfe eines Items entspricht dabei der Korrelation der Messwerte des jeweiligen Items mit dem Summenscore der restlichen Items der Skala. Entsprechend der vorgenommenen Umpolung der Items können nachfolgend ausschließlich positive Trennschärfen entstehen. Hohe Itemtrennschärfen zeigen an, dass das jeweilige Item von Befragten, die einen hohen Testwert erreichen, in Richtung der Zieldimension beantwortet wird, während Befragte mit niedrigen Testwerten entgegengesetzt der Zieldimension antworten. Bei einer hohen Trennschärfe kann man dementsprechend davon ausgehen, dass das jeweilige Item sehr ähnlich differenziert wie die Gesamtskala. (Bühner (2011), S. 171ff; Fisseni (1997), S. 49ff; Moosbrugger/Kelava (2012), S. 84ff)

Ziel der Itemselektion anhand von Itemschwierigkeit und Trennschärfe ist die Erstellung eines homogenen Tests. Das bedeutet: Die resultierende Skala soll solche Items enthalten, die möglichst hoch miteinander korrelieren und auf inhaltlicher Ebene die Zieldimension gut repräsentieren. Aus diesem Grund werden vornehmlich solche Items in die Vertrauensskala aufgenommen, die eine hohe Trennschärfe bei mittlerer Itemschwierigkeit und hoher Itemvarianz aufweisen. Um auch zwischen Befragten mit extremen Merkmalsausprägungen differenzieren zu können, werden zusätzlich auch Items mit niedrigen bzw. hohen Schwierigkeitsindizes beibehalten, wenn Sie eine ausreichend hohe Trennschärfe aufweisen. Es sollen somit sowohl psychometrisch leichte wie schwere Items in der Skala vertreten sein, um eine differenzierte Messung auch in den Randbereichen des Merkmals „Generalisiertes Vertrauen“ zu ermöglichen. (Bühner (2011), S. 223, S. 248f; Fisseni (1997), S. 55, S. 60; Lienert/Raatz (1998), S. 115; Moosbrugger/Kelava (2012), S. 87)

6. Pilotstudie

Tab. 6.4: Itemkennwerte der Rohskala „Generalisiertes Vertrauen“ (N = 220)

Item	P ^a	P _m ^b	σ_i^c	Sch _i ^d	Ex _i ^e	P _{mVar} ^f	$r_{1(x-y)}^g$	Sel ^h	SK ⁱ
v2.1	3,7409	,685	,93201	-,212	-,698	,524	,274**	0,295	0,147
v2.2	2,0227	,256	,95297	,849	,463	,122	,289**	0,331	0,152
v2.3	3,2909	,573	,89998	-,267	-,141	,378	,231**	0,233	0,128
v2.4	3,8273	,707	,83703	-,891	1,376	,543	,207**	0,227	0,124
v2.5	2,7227	,431	1,08136	,090	-,691	,258	,229**	0,231	0,106
v2.6	2,0818	,270	,81237	,725	,887	,114	,384**	0,432	0,236
v2.7	2,2182	,305	,86939	,529	,133	,140	,319**	0,347	0,183
v2.8	2,9182	,480	,98974	-,177	-,561	,291	,314**	0,314	0,159
v2.9	3,2955	,574	,98341	-,622	-,361	,389	,191**	0,193	0,097
v2.10	2,9091	,477	1,01626	-,080	-,533	,292	,307**	0,307	0,151
v2.11	3,0636	,516	,80837	-,169	-,256	,307	,449**	0,449	0,278
v2.12	3,6727	,668	1,09885	-,634	-,465	,522	,258**	0,274	0,117
v2.13	3,0682	,517	,91656	-,136	-,368	,320	,439**	0,439	0,239
v2.14	2,2773	,319	1,00702	,476	-,604	,165	,307**	0,329	0,152
v2.15	3,5682	,642	1,11850	-,390	-,702	,490	,440**	0,459	0,197
v2.16	3,0455	,511	,91548	-,415	-,381	,314	,260**	0,260	0,142
v2.17	2,5545	,389	1,05628	,125	-1,061	,220	,324**	0,332	0,153
v2.18	2,1364	,284	1,06396	,712	-,175	,151	,023	0,026	0,011
v2.19	2,1455	,286	1,19997	,821	-,342	,172	,183**	0,202	0,076
v2.20	3,3273	,582	,98953	-,553	,002	,399	,168**	0,170	0,085
v2.21	4,0227	,756	,91882	-,758	,304	,624	,089	0,104	0,048
v2.22	4,2455	,811	,90343	-1,405	2,111	,709	,143*	0,183	0,079
v2.23	3,0545	,514	,99622	-,082	-,459	,326	,340**	0,340	0,171
v2.24	3,3955	,599	,92306	-,129	-,615	,412	,381**	0,389	0,206
v2.25	2,8909	,473	1,07987	,021	-,839	,296	,258**	0,258	0,119
v2.26	3,3409	,585	1,07998	-,340	-,562	,415	,167**	0,169	0,077
v2.27	3,2818	,570	1,20222	-,288	-,876	,415	,156*	0,158	0,065
v2.28	2,8182	,455	,93810	-,198	-,547	,261	,261**	0,262	0,139
v2.29	3,8273	,707	,97324	-,667	-,089	,559	,206**	0,226	0,106
v2.30	2,2091	,302	,90756	,757	,552	,143	,165**	0,180	0,091
v2.31	3,0636	,516	1,16476	,050	-,781	,351	,008	0,008	0,003
v2.32	2,2136	,303	,90397	,500	-,097	,143	,447**	0,486	0,247
v2.33	2,5545	,389	1,02111	,357	-,682	,216	,283**	0,290	0,139
v2.34	3,1273	,532	1,25774	-,159	-1,077	,381	,147*	0,147	0,058
v2.35	3,4045	,601	1,18398	-,345	-,843	,449	,030	0,031	0,013
v2.36	4,2182	,805	,88500	-1,160	1,341	,696	,196**	0,247	0,111
v2.37	3,7500	,688	,85808	-,457	-,107	,518	,465**	0,502	0,271
v2.38	2,8273	,457	,94467	,188	-,716	,264	,061	0,061	0,032
v2.39	3,4045	,601	,77929	-,613	,501	,399	,350**	0,357	0,225
v2.40	3,2227	,556	1,05572	-,245	-,467	,378	,361**	0,363	0,171
v2.41	2,5091	,377	1,04024	,283	-,596	,210	,450**	0,464	0,216
v2.42	2,8636	,466	1,02460	,251	-,434	,282	,493**	0,494	0,241
ϕ^j	3,0508	,513	,99025	---	---	,347	,263	---	---

Sortiert nach Itembezeichnung

^a Itemschwierigkeit (Mittelwert)^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)⁴²^c Itemstreuung (Standardabweichung)^d Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)^e Exzess bzw. Kurtosis (Ex_i < 0 → breitipflige Verteilung; Ex_i > 0 → schmalipflige Verteilung) (Bühner (2006), S. 88f)^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)^h Selektionskennwert nach Lienert/Raatz (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)ⁱ Selektionskennwert nach Bühner (Bühner (2006), S. 99f)^j Durchschnittswert* $\alpha < .05$ (einseitig), ** $\alpha < .01$ (einseitig)

(Quelle: eigene Berechnungen)

⁴² Die Berechnung der mehrstufigen Schwierigkeitsindizes ist nur dann möglich, wenn die Werte der Antwortskala auf einer kontinuierlichen Skala von 0; 1; 2 ... n liegen. Um eine Vergleichbarkeit der Schwierigkeitsindizes zu gewährleisten, müssen zusätzlich alle Items in Richtung der Zieldimension gepolt werden. Aus diesem Grund wurde zur die Bestimmung des Schwierigkeitsindex für mehrstufige Items eine Umcodierung der Skalenwerte vorgenommen: 1→0 (Stimme nicht zu), 2→1 (Stimme eher nicht zu), 3→2 (Stimme teils zu, stimme teils nicht zu), 4→3 (Stimme eher zu) und 5→4 (Stimme zu). Außerdem wurde für die Items v2.6, v2.7, v2.9, v2.10, v2.14, v2.15, v2.17, v2.18, v2.19, v2.22, v2.23, v2.25, v2.27, v2.28, v2.29, v2.30, v2.32, v2.33, v2.35, v2.36, v2.41 und v2.42 die Antwortskala umgepolt. Auf diese Weise gilt für alle Items der Rohskala, dass ein hoher Skalenwert einer starken Ausprägung der Zieldimension „Generalisiertes Vertrauen“ entspricht. (Bühner (2011), S. 217ff; Dahl (1971))

6.2.1.1.1 Schwierigkeitsanalyse

Zunächst ist festzuhalten, dass bei allen Items, die in der Rohskala „Generalisiertes Vertrauen“ zusammengefasst sind, die gesamte Breite der Antwortskala ausgenutzt wurde. So ist jede Antwortkategorie mit mindestens einem Fall besetzt, was darauf hinweist, dass das angewandte Antwortformat ausreichend stark ausdifferenziert ist. (Bühner (2011), S. 233) Mit Ausnahme von Item v2.34 und v2.12, welche bimodal verteilt sind, weisen alle verbleibenden Items unimodale Verteilungen auf. Die meisten Items wurden somit von allen Befragten in ähnlicher bzw. gleicher Art und Weise verstanden und es kann weitestgehend von einer homogenen Stichprobe ausgegangen werden. Die Items v2.34 und v2.12 können allerdings bereits an dieser Stelle von der weiteren Analyse ausgeschlossen und aus der Skala entfernt werden. (Bühner (2011), S. 256)

Die Itemschwierigkeit variiert in einem Bereich von $P_i=2,0227$ bis $P_i=4,2455$, wobei die Mehrzahl der Items (ca. 57%) eine Schwierigkeit aufweist, die oberhalb des theoretischen Skalenmittelpunktes von 3 angesiedelt ist. Die durchschnittliche Itemschwierigkeit beträgt $P_i=3,0508$ und ist damit vom theoretischen Skalenmittelpunkt leicht nach rechts verschoben, wobei die Itemschwierigkeiten der einzelnen Items im Mittel um rund einen Skalenpunkt von diesem Durchschnittswert abweichen.

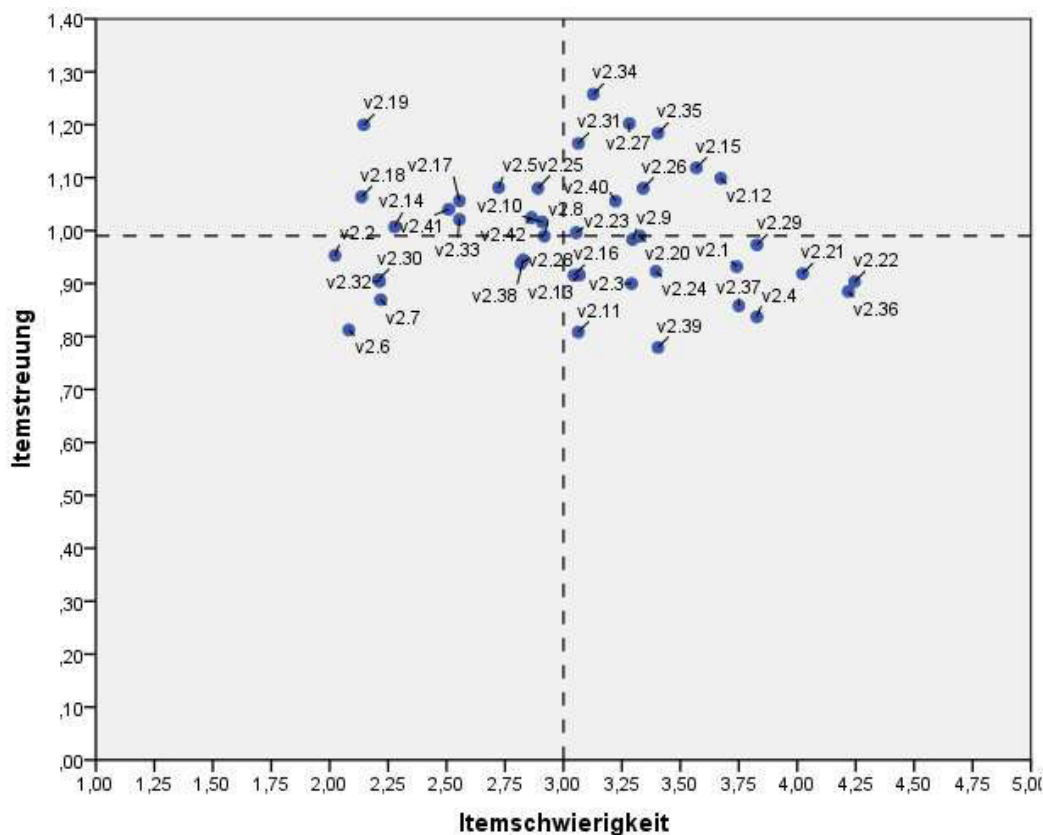


Abb. 6.1: Zusammenhang zwischen Itemschwierigkeit (arithmetisches Mittel $[P_i]$) und Itemstreuung (Standardabweichung) für alle Items der Rohskala „Generalisiertes Vertrauen“

Zwar enthält die Rohskala vorwiegend Items, bei denen die Befragten der Zieldimension „Generalisiertes Vertrauen“ eher zustimmend gegenüberstanden, betrachtet man allerdings zusätzlich den *Schwierigkeitsindex nach DAHL*, so kommt man zu dem Schluss, dass die Skala vornehmlich Items mit psychometrisch mittlerer Schwierigkeit beinhaltet (ca.95%). Keines der

6. Pilotstudie

Items weist eine Itemschwierigkeit im niedrigen Bereich auf und lediglich zwei Items besitzen hohe Schwierigkeiten (ca. 5%). Es lässt sich somit allenfalls ein leichter Deckeneffekt beobachten, was bedeutet, dass die Befragten auf den vorliegenden Items tendenziell eher hohe Vertrauenswerte erreichen. (Bühner (2006), S. 98)

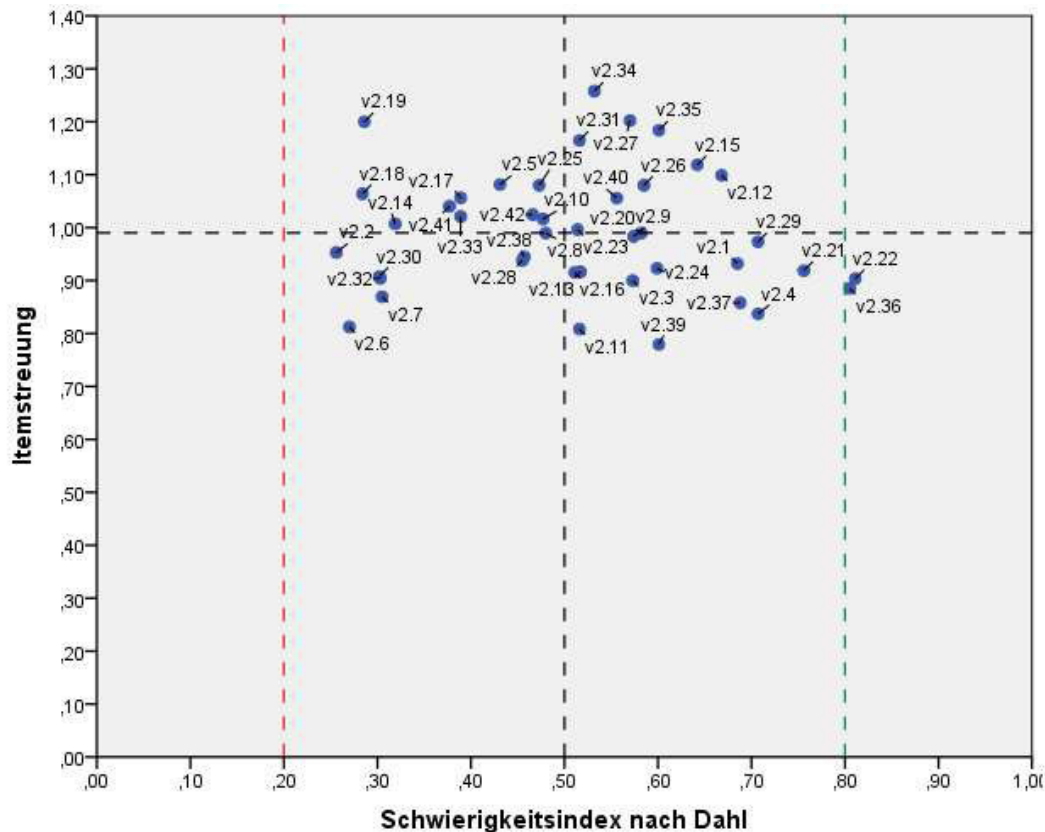


Abb. 6.2: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m]) und Itemstreuung (Standardabweichung) für alle Items der Rohskala „Generalisiertes Vertrauen“; ($P > .80$ = Niedrig; $P = .80 - .20$ = Mittel; $P < .20$ = Hoch) (Bühner (2011), S. 81)

Bezieht man die Itemstreuungen in die Berechnung des Schwierigkeitsindexes ein (*Schwierigkeitsindex nach FISSENI*), so verschieben sich sämtliche Itemschwierigkeiten nach links. Keines der Items erreicht nun mehr eine hohe Schwierigkeit, dafür weisen nun ca. 19% der Items niedrige Schwierigkeitswerte auf. Die verbleibenden Items (ca. 81%) besitzen mittlere Itemschwierigkeiten. Insgesamt liegen nun nur noch ca. 19% der Items rechts vom theoretischen Skalenmittelpunkt und rund 81% links von diesem. Damit kehrt sich das Bild von einem leichten Deckeneffekt radikal in das eines starken Bodeneffektes um. In der Tendenz erreichen die Befragten somit eher niedrige Vertrauenswerte auf den Items der Skala. (Bühner (2006), S. 98; Bühner (2011), S. 81, S. 242; Fisseni (1997), S. 124)

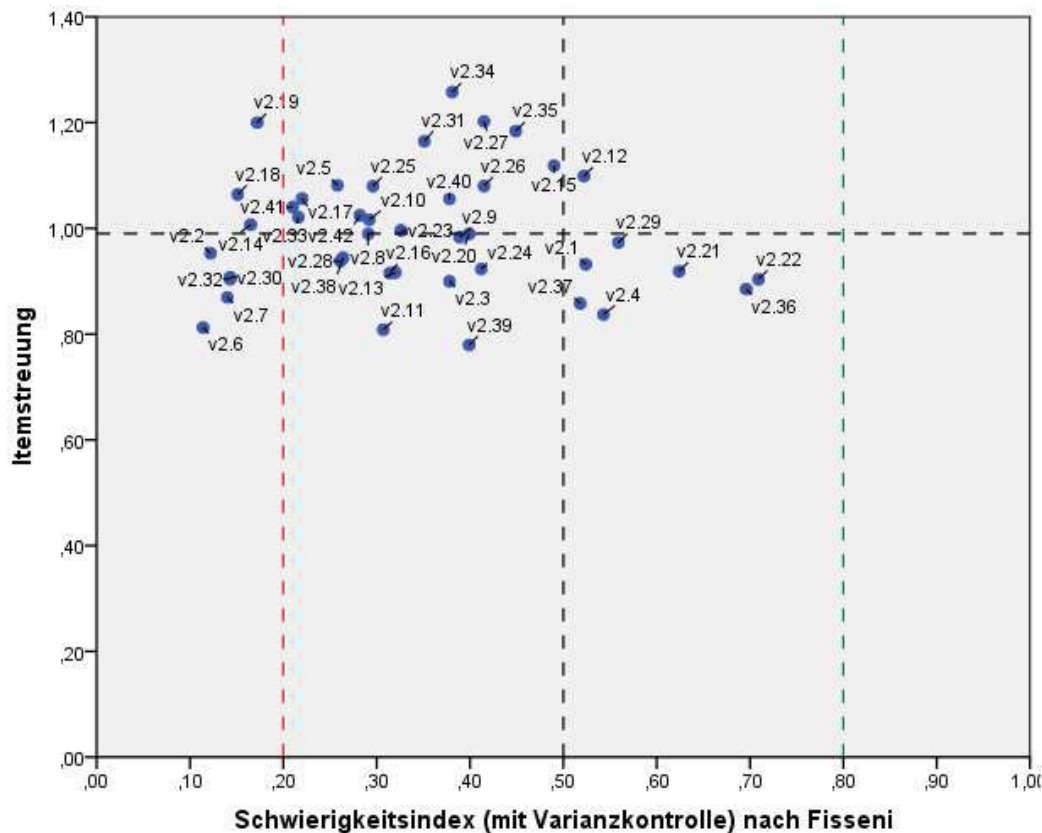


Abb. 6.3: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach FISSENI [P_{mVar}]) und Itemstreuung (Standardabweichung) für alle Items der Rohskala „Generalisiertes Vertrauen“; ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) (Bühner (2011), S. 81)

6.2.1.1.1 Items mit mittlerer Schwierigkeit und überdurchschnittlicher Streuung

Nachfolgend sollen nun die Schwierigkeiten, Standardabweichungen und Häufigkeitsverteilungen der einzelnen Items betrachtet werden, um jeweils zu entscheiden, ob das Item ausreichend gut zwischen verschiedenen Ausprägungen Generalisierten Vertrauens differenzieren kann. Hierbei sind insbesondere solche Items auszuwählen, die eine mittlere Itemschwierigkeit aufweisen, die zusätzlich möglichst nahe am theoretischen Skalenmittelpunkt von 3 liegt. FISSENI schlägt im Rahmen der Erstellung eines homogenen Tests ein Schwierigkeitsintervall von $P_m = .40-.60$ vor. (Fisseni (1997), S. 60) Solche Items ermöglichen mehr Unterscheidungen zwischen Befragten als Items mit extremen Schwierigkeiten, die weit links oder rechts vom theoretischen Skalenmittelpunkt liegen. Mit anderen Worten: Items mit mittleren Schwierigkeiten trennen besser zwischen Personen mit niedrigen und hohen Merkmalsausprägungen. (Bühner (2006), S. 98f; Lienert/Raatz (1998), S. 105) Außerdem sind für Items mit mittleren Schwierigkeiten höhere Trennschärfen und auch höhere Reliabilitäten zu erwarten. (Bühner (2006), S. 99)

Des Weiteren sollten die Items eine möglichst hohe Streuung aufweisen. Die Itemvarianz kann dabei als die Differenzierungsfähigkeit eines Items hinsichtlich der Zieldimension verstanden werden und steht in direkter Verbindung mit der Itemschwierigkeit. So weisen in der Regel Items mit mittleren Schwierigkeiten höhere Streuungen auf als Items mit extrem niedrigen oder

6. Pilotstudie

extrem hohen Schwierigkeiten. Eine breite Streuung der Antworten der Befragten über eine Skala ist demzufolge als ein Indiz für eine stark ausgeprägte Differenzierungsfähigkeit zu werten. Je ähnlicher die Befragten dagegen auf einem Item antworten, desto weniger gut kann ein Item verschiedene Merkmalsausprägungen voneinander unterscheiden. (Bühner (2011), S. 220ff, S. 258; Moosbrugger/Kelava (2012), S. 81ff). Als Schwellenwert für eine hohe Streuung wurde in dieser Analyse der Mittelwert der Standardabweichungen aller Items gewählt, die in der vorläufigen Skala „Generalisiertes Vertrauen“ enthalten sind. Dieser entspricht $\sigma \approx 0,99025$.

Ferner sollte das Item eine geringe Schiefe und eine breitgipflige Verteilung anzeigen bzw. möglichst symmetrisch sein. (Bühner (2011), S. 228ff; Lienert/Raatz (1998), S. 109)

Treten diese Kriterien gemeinsam auf, so kann davon ausgegangen werden, dass sich die Antworten der Befragten breit über alle Antwortkategorien verteilen und das Item Unterschiede zwischen hohem und niedrigem Generalisiertem Vertrauen gut abbilden kann.

In einem ersten Analyseschritt werden zunächst diejenigen Items ausgewählt, deren Schwierigkeiten sowohl auf dem *Schwierigkeitsindex nach DAHL* als auch dem *Schwierigkeitsindex nach FISSENI* im mittleren Schwierigkeitsbereich ($P_m = .20-.80$ bzw. $P_{mVar} = .20-.80$) liegen. Items mit extremen Schwierigkeiten werden auf diese Weise ausgeschlossen und es bleiben nur solche Items erhalten, bei denen es den Befragten weder zu leicht noch zu schwer fällt, in Schlüsselrichtung zuzustimmen. (Bühner (2011), S. 220; Moosbrugger/Kelava (2012), S. 81)

In einem zweiten Analyseschritt werden Items mit unterdurchschnittlicher Streuung ($\sigma_i < .99025$) aussortiert. Auf diese Weise bleiben nur Items erhalten, deren Werte verhältnismäßig breit über die Antwortskala streuen und demzufolge eine stark ausgeprägte Differenzierungsfähigkeit besitzen.

In der Folge dieser beiden Analyseschritte werden ausschließlich diejenigen Items selektiert, die bei einer mittleren Itemschwierigkeit überdurchschnittliche Standardabweichungen aufweisen. Unter Berücksichtigung der bereits wegen einer bimodalen Häufigkeitsverteilung aussortierten Items v2.12 und v2.34 resultieren dann die folgenden Items:

Tab. 6.5: Itemkennwerte der Rohskala: mittlere Schwierigkeit; überdurchschnittliche Streuung (N = 220)

Item	P_i^a	P_m^b	σ_i^c	Sch_i^d	Ex_i^e	P_{mVar}^f	$r_1(x-i)^g$	Sel^h
v2.15	3,5682	,642	1,11850	-,390	-,702	,490	,440**	0,459
v2.35	3,4045	,601	1,18398	-,345	-,843	,449	,030	0,031
v2.26	3,3409	,585	1,07998	-,340	-,562	,415	,167**	0,169
v2.27	3,2818	,570	1,20222	-,288	-,876	,415	,156*	0,158
v2.40	3,2227	,556	1,05572	-,245	-,467	,378	,361**	0,363
v2.31	3,0636	,516	1,16476	,050	-,781	,351	,008	0,008
v2.23	3,0545	,514	,99622	-,082	-,459	,326	,340**	0,340
v2.10	2,9091	,477	1,01626	-,080	-,533	,292	,307**	0,307
v2.25	2,8909	,473	1,07987	,021	-,839	,296	,258**	0,258
v2.42	2,8636	,466	1,02460	,251	-,434	,282	,493**	0,494
v2.5	2,7227	,431	1,08136	,090	-,691	,258	,229**	0,231
v2.17	2,5545	,389	1,05628	,125	-1,061	,220	,324**	0,332
v2.33	2,5545	,389	1,02111	,357	-,682	,216	,283**	0,290
v2.41	2,5091	,377	1,04024	,283	-,596	,210	,450**	0,464

Sortiert nach P_i

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)

^e Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach Lienert/Raatz (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

* $\alpha < .05$ (einseitig), ** $\alpha < .01$ (einseitig)

(Quelle: eigene Berechnungen)

Das Streudiagramm zeigt, dass sich die selektierten Items zumindest auf dem *Schwierigkeitsindex nach DAHL* in etwa im Rahmen des von FISSENI für homogene Tests aufgespannten Intervalls von $P_m = .40-.60$ bewegen. So stellt v2.41 das vergleichsweise psychometrisch leichteste Item dar. Es erreicht eine Schwierigkeit von $P_m = .377$ bzw. $P_{mVar} = .242$ während sich mit v2.15 und $P_m = .642$ bzw. $P_{mVar} = .490$ das vergleichsweise psychometrisch schwerste Item ergibt.

Der Einbezug der Itemvarianzen bei der Berechnung des *Schwierigkeitsindex nach FISSENI* sorgt indes dafür, dass sich die Schwierigkeitswerte für alle Items auf der Bewertungsskala nach links verschieben. Als Erklärung für diesen Effekt kann herangezogen werden, dass die Standardabweichungen der ausgewählten Items zwar relativ zu den restlichen Items als hoch zu bewerten sind, setzt man allerdings absolute Maßstäbe an, so scheinen diese trotz allem eher niedrig zu sein. So würde die Schwierigkeit der Items auf diesem Index bei einer maximalen Varianz gegen $P_m = .50$ tendieren. Der *Schwierigkeitsindex nach FISSENI* verschiebt somit Items mit eher niedrigen bzw. hohen Schwierigkeiten und eher niedrigen Varianzen mehr in Richtung des jeweiligen Schwierigkeitswertes. Dabei spielen insbesondere die Extremwerte eine große Rolle. So verschiebt sich die Schwierigkeit mehr Richtung $P_m = 1$, wenn viele Befragte den maximalen Item-Score erreichen, während sich die Schwierigkeit mehr Richtung $P_m = 0$ verschiebt, wenn mehr Befragte den minimalen Item-Score erreichen. Gleichzeitig weist die Mehrzahl der Items leicht linkssteile Verteilungen auf, sodass sich die Fälle eher links vom theoretischen Skalenmittelwert sammeln und sich ein sogenannter Bodeneffekt einstellt. (Fisseni (1997), S. 45)

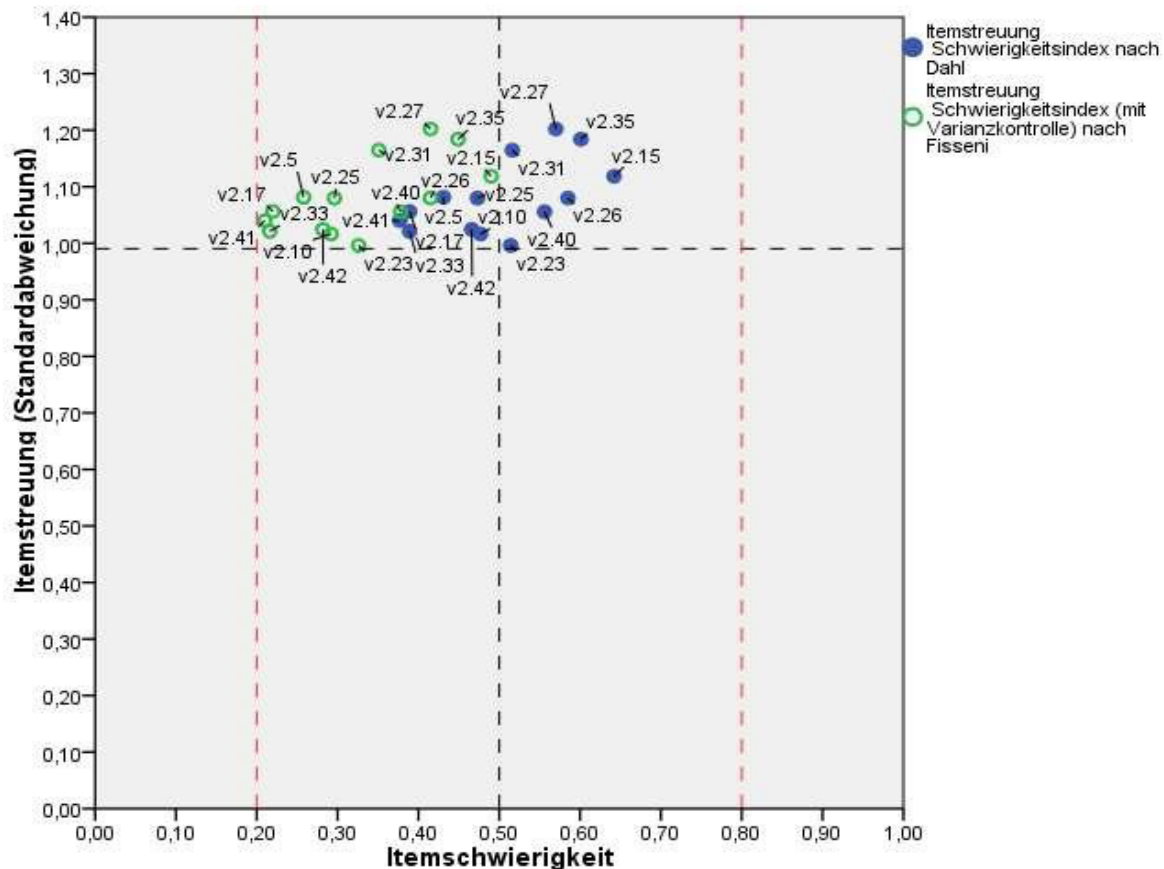


Abb. 6.4: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemstreuung (Standardabweichung) für Items der Rohskala „Generalisiertes Vertrauen“ mit mittlerer Schwierigkeit; ($P_m = .20-.80$ bzw. $P_{mVar} = .20-.80$) und überdurchschnittlicher Streuung (hier: $\sigma \geq 0,99025$); ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) (Bühner (2011), S. 81)

Betrachtet man die Histogramme der Items mit günstigen Schwierigkeiten und überdurchschnittlichen Streuungen, so zeigen sich breitipflige und vornehmlich symmetrische bzw. nahezu symmetrische Verteilungen. Je weiter sich die Schwierigkeiten nach links bzw. rechts vom theoretischen Skalenmittelpunkt von 3 entfernen, desto stärker ist die Schiefe ausgeprägt. Nach rechts verschobene Items weisen dementsprechend eher rechtssteile Verteilungen und nach links verschobene Items eher linkssteile Verteilungen auf. In allen Fällen ergeben sich durch die überdurchschnittlichen Streuungen allerdings breitipflige Verteilungen, sodass auch auf grafischem Wege gezeigt werden kann, dass sich die Antworten der Befragten mehr oder weniger gleichmäßig über die gesamte Antwortskala verteilen. Es ist dementsprechend anzunehmen, dass die Items v2.5, v2.10, v2.15, v2.17, v2.23, v2.25 bis v2.27, v2.31, v2.33, v2.35 sowie v2.40 bis v2.42 besonders gut geeignet sind, um zwischen Personen mit schwach bzw. stark ausgeprägtem Generalisiertem Vertrauen zu unterscheiden.⁴³

⁴³ Anhang VIII: Häufigkeitsverteilungen (Pilotstudie), VIII.a Rohskala „Generalisiertes Vertrauen“

6.2.1.1.2 Items mit mittlerer Schwierigkeit und unterdurchschnittlicher Streuung

Eine Iteration der Analyseschritte eins und zwei mit umgekehrten Voraussetzungen in Schritt zwei ergibt diejenigen Items, welche sowohl hinsichtlich der *Schwierigkeitsindizes nach DAHL* und *FISSENI* mittlere Schwierigkeiten, aber unterdurchschnittliche Itemstreuungen aufweisen. Für die auf diese Weise selektierten Items ist anzunehmen, dass sie zwar theoretisch mehr Unterscheidungen zwischen den Befragten ermöglichen sollten, tatsächlich aber nur über eine eher geringe Differenzierungsfähigkeit verfügen. (Bühner (2006), S. 98f; Lienert/Raatz (1998), S. 105) Es ergeben sich dann die folgenden Items:

Tab. 6.6: Itemkennwerte der Rohskala: mittlere Schwierigkeit; unterdurchschnittliche Streuung (N = 220)

Item	P _i ^a	P _m ^b	σ _i ^c	Sch _i ^d	Ex _i ^e	P _{mVar} ^f	r _{i (K-i)} ^g	Sel ^h
v2.21	4,0227	,756	,91882	-,758	,304	,624	,089	0,104
v2.4	3,8273	,707	,83703	-,891	1,376	,543	,207**	0,227
v2.29	3,8273	,707	,97324	-,667	-,089	,559	,206**	0,226
v2.37	3,7500	,688	,85808	-,457	-,107	,518	,465**	0,502
v2.1	3,7409	,685	,93201	-,212	-,698	,524	,274**	0,295
v2.39	3,4045	,601	,77929	-,613	,501	,399	,350**	0,357
v2.24	3,3955	,599	,92306	-,129	-,615	,412	,381**	0,389
v2.20	3,3273	,582	,98953	-,553	,002	,399	,168**	0,170
v2.9	3,2955	,574	,98341	-,622	-,361	,389	,191**	0,193
v2.3	3,2909	,573	,89998	-,267	-,141	,378	,231**	0,233
v2.13	3,0682	,517	,91656	-,136	-,368	,320	,439**	0,439
v2.11	3,0636	,516	,80837	-,169	-,256	,307	,449**	0,449
v2.16	3,0455	,511	,91548	-,415	-,381	,314	,260**	0,260
v2.8	2,9182	,480	,98974	-,177	-,561	,291	,314**	0,314
v2.38	2,8273	,457	,94467	,188	-,716	,264	,061	0,061
v2.28	2,8182	,455	,93810	-,198	-,547	,261	,261**	0,262

Sortiert nach P_i

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)

^e Exzess bzw. Kurtosis (Ex_i < 0 → breitipflige Verteilung; Ex_i > 0 → schmalipflige Verteilung) (Bühner (2006), S. 88f)

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach Lienert/Raatz (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

*α < .05 (einseitig), **α < .01 (einseitig)

(Quelle: eigene Berechnungen)

Jedes der nach den obengenannten Vorgaben ausgewählten Items liegt innerhalb des Bereichs für mittlere Itemschwierigkeiten (P_m = .20-.80 bzw. P_{mVar} = .20-.80). Allerdings überschreiten bzw. unterschreiten einige Items, je nach angelegtem Schwierigkeitsindex, die von FISSENI für homogene Tests vorgeschlagenen Grenzwerte. Dabei erreichen die Items v2.21 mit P_m = .756 bzw. P_{mVar} = .624 und Item v2.28 mit P_m = .455 bzw. P_{mVar} = .261 die extremsten Schwierigkeiten.

Nach DAHL liegen die Schwierigkeiten der meisten Items rechts vom theoretischen Skalenmittelpunkt. Es ist daher zu erwarten, dass es den Befragten in der Tendenz eher leichter fiel, den Items zu zustimmen und somit in Schlüsselrichtung der Skala zu antworten. Nach FISSENI bzw. unter Kontrolle der Itemvarianzen ergibt sich ein entgegengesetztes Bild. Die Itemschwierigkeiten ordnen sich nun vornehmlich links vom theoretischen Skalenmittelpunkt an. Hiernach müsste es den Befragten in der Tendenz eher schwerer gefallen sein, den Items zu zustimmen bzw. in Richtung der Zieldimension „Generalisiertes Vertrauen“ zu antworten.

Dieser Effekt lässt sich vornehmlich durch das Verhältnis der Häufigkeiten der Randkategorien zueinander und die Itemstreuungen erklären. Entsprechend der Spezifikationen des Schwierigkeitsindex nach FISSENI sollte sich der Index in Richtung P_m = 1 bewegen, wenn viele

Befragte den maximalen Score erreichen, und er sollte sich in Richtung $P_m = 0$ bewegen, wenn viele Befragte den minimalen Score erreichen. Ausgehend von diesen Vorgaben und der unterdurchschnittlichen Itemvarianzen hätten sich rechtssteile Items weiter nach rechts und linkssteile Items weiter nach links verschieben müssen. Gleichzeitig sollte sich die Schwierigkeit in Richtung $P_m = .50$ bewegen, wenn die Itemvarianz maximal wird. Entsprechend sollten Items mit vergleichsweise hohen Standardabweichungen in Richtung der theoretischen Skalenmitte verschoben werden. (Fisseni (1997), S. 45)

Für die Items v2.1, v2.3, v2.4, v2.20, v2.21, v2.24, v2.29, v2.37 und v2.39 lässt sich feststellen, dass die linke Randkategorie deutlich schwächer besetzt ist als die rechte Randkategorie. Die Items weisen dementsprechend eine Tendenz in Richtung $P_m = 1$ auf. Gleichzeitig sind aber die Itemstreuungen noch ausreichend hoch, sodass sich ebenfalls eine Tendenz zur Mitte ergibt. Beide Tendenzen kompensieren sich bis zu einem gewissen Grad. Je stärker diese Kompensation, desto weniger stark verschiebt sich der *Schwierigkeitsindex nach FISSENI* relativ zum *Schwierigkeitsindex nach DAHL*.

Für die Items v2.8, v2.11, v2.28 und v2.38 lässt sich ein umgekehrtes Verhältnis der Randkategorien beobachten. Hier ist die linke Randkategorie stärker mit Fällen besetzt als die rechte Randkategorie, was eine Tendenz in Richtung $P_m = 0$ ergibt. Auch hier wirkt die Itemstreuung diesem Effekt entgegen.

Für die Items v2.9 und v2.13 ergibt sich ein Gleichgewicht der Fälle auf den Randkategorien. Im Fall von v2.9 ergibt eine in Relation zu v2.13 eine höhere Streuung, was eine stärkere Tendenz zur Mitte erwarten lässt. Gleichzeitig weist v2.9 eine stärker rechtssteile Verteilung auf als v2.13. In der Summe bedeutet dies, dass v2.9 weniger nach links verschoben wird als v2.13.

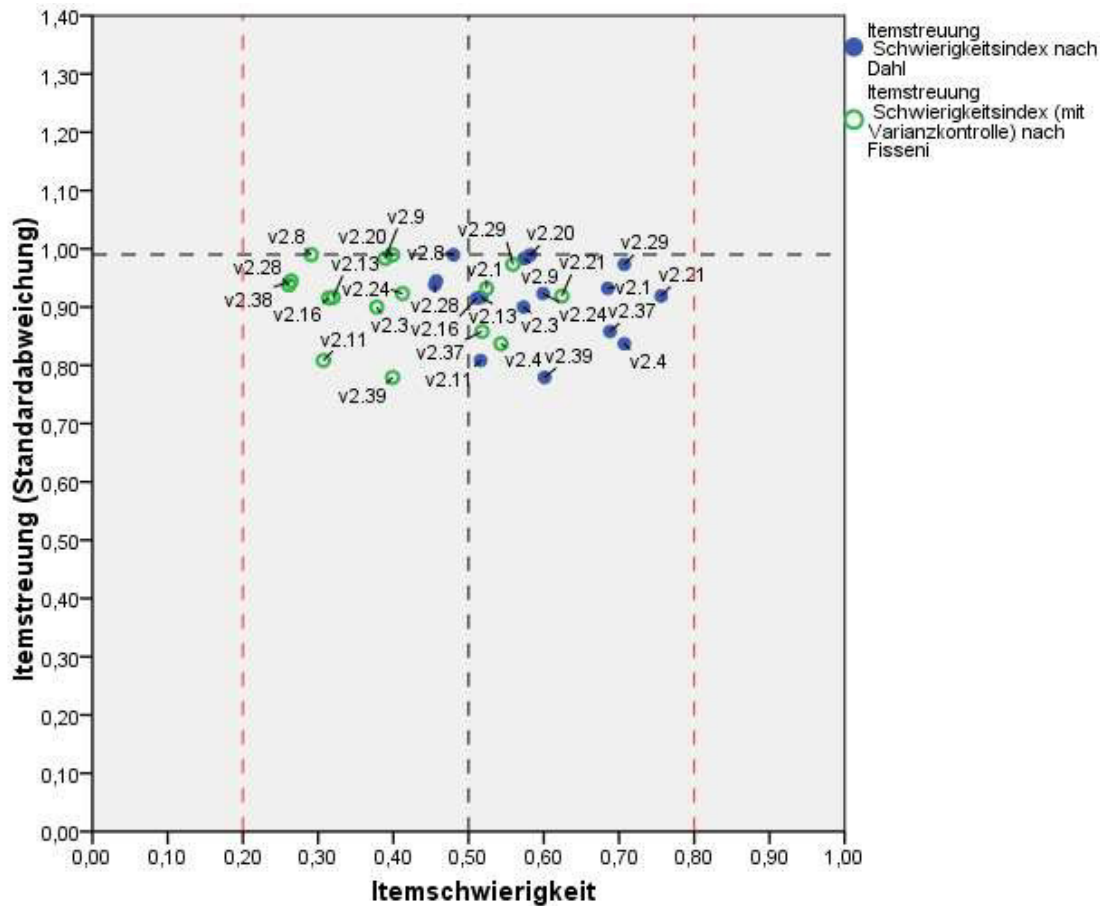


Abb. 6.5: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemstreuung (Standardabweichung) für Items der Rohskala „Generalisiertes Vertrauen“ mit mittlerer Schwierigkeit ($P_m = .20-.80$ bzw. $P_{mVar} = .20-.80$) und unterdurchschnittlicher Streuung (hier: $\sigma < 0,99025$); ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) (Bühner (2011), S. 81)

Aus der grafischen Analyse der Items geht hervor, dass es sich in der Mehrzahl um rechtssteile und eher breitgipflige Verteilungen handelt. Lediglich fünf der Items entsprechen nicht diesem Muster. So sind v2.4, v2.20, v2.21 und v2.39 rechtssteil und eher schmalgipflig verteilt, während v2.38 eher linkssteil und breitgipflig verteilt ist.

Aufgrund der ausgeprägten Schiefe und der unterdurchschnittlichen Varianz ist für die Items v2.21, v2.4, v2.29, v2.37, v2.1, v2.39, v2.24, v2.20, v2.9 zu erwarten, dass das wahre Ausmaß Generalisierten Vertrauens eher überschätzt wird. Die Schwierigkeiten der genannten Items sind aus der Perspektive der theoretischen Skalenmitte stark nach rechts verschoben und streuen eher über die rechten Randkategorien als über die gesamte Antwortskala. Da es den Befragten augenscheinlich eher leicht fiel, den Statements zu zustimmen, scheinen diese Items weniger gut geeignet zu sein, „Generalisiertes Vertrauen“ differenziert zu messen.⁴⁴

Die übrigen Items weisen zwar vergleichsweise günstige Itemschwierigkeiten auf, allerdings ergibt sich infolge der unterdurchschnittlichen Itemstreuungen nur eine eher gering ausgeprägte Differenzierungsfähigkeit. Aus den zugehörigen Histogrammen geht hervor, dass die Items zwar im mittleren Bereich der Antwortskala noch ausreichend gut differenzieren, extreme Ausprägungen Generalisierten Vertrauens aber weit weniger gut abgebildet werden

⁴⁴ Anhang VIII: Häufigkeitsverteilungen (Pilotstudie), VIII.a Rohskala „Generalisiertes Vertrauen“

6. Pilotstudie

können. Auch diese Items scheinen demnach weniger gut für eine differenzierte Darstellung Generalisierten Vertrauens geeignet.⁴⁵

6.2.1.1.3 Items mit extrem niedriger Schwierigkeit und variabler Streuung

Wiederholt man Analyseschritt eins in der Art, dass nur diejenigen Items ausgewählt werden, deren Schwierigkeiten entweder auf dem *Schwierigkeitsindex nach DAHL* oder auf dem *Schwierigkeitsindex nach FISSENI* im niedrigen Schwierigkeitsbereich ($P_m < .20$ bzw. $P_{mVar} < .20$) liegen, so werden Items mit mittleren und extrem hohen Schwierigkeiten ausgeschlossen. Es bleiben nur solche Statements erhalten, bei denen es den Befragten besonders schwerfiel, in Schlüsselrichtung zu antworten bzw. dem Item zu zustimmen (sog. psychometrisch schwere Items). (Bühner (2011), S. 220; Moosbrugger/Kelava (2012), S. 81) Es resultieren die nachfolgenden Items:

Tab. 6.7: Itemkennwerte der Rohskala: extrem niedrige Schwierigkeit; variable Streuung (N = 220)

Item	P_i^a	P_m^b	σ_i^c	Sch_i^d	Ex_i^e	P_{mVar}^f	$r_1(x-i)^g$	Sel^h
v2.14	2,2773	,319	1,00702	,476	-,604	,165	,307**	0,329
v2.7	2,2182	,305	,86939	,529	,133	,140	,319**	0,347
v2.32	2,2136	,303	,90397	,500	-,097	,143	,447**	0,486
v2.30	2,2091	,302	,90756	,757	,552	,143	,165**	0,180
v2.19	2,1455	,286	1,19997	,821	-,342	,172	,183**	0,202
v2.18	2,1364	,284	1,06396	,712	-,175	,151	,023	0,026
v2.6	2,0818	,270	,81237	,725	,887	,114	,384**	0,432
v2.2	2,0227	,256	,95297	,849	,463	,122	,289**	0,331

Sortiert nach P_i

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)

^e Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach Lienert/Raatz (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

* $\alpha < .05$ (einseitig), ** $\alpha < .01$ (einseitig)

(Quelle: eigene Berechnungen)

Die ausgewählten Items erreichen jeweils Schwierigkeiten, die auf dem Schwierigkeitsindex nach DAHL jenseits des von FISSENI aufgespannten Intervalls hinsichtlich für homogene Tests günstigen Itemschwierigkeiten von $P_m = .40-.60$ liegen. Auf dem Schwierigkeitsindex nach FISSENI erreichen diese Items lediglich Werte im niedrigen Schwierigkeitsbereich, sodass die Items eher psychometrisch schwere Statements darstellen und es den Befragten somit eher schwer fiel zuzustimmen. Psychometrisch besonders leicht war Item v2.2 mit $P_m = .256$ bzw. $P_{mVar} = .122$. Diesem Item haben im Skalenvergleich die wenigsten Personen entsprechend der Zieldimension zugestimmt.

Die Linksverschiebung der Schwierigkeiten durch den Index nach FISSENI ist darauf zurückzuführen, dass es sich durchgängig um linkssteile Verteilungen handelt und sich somit die Mehrzahl der Fälle links vom theoretischen Skalenmittelpunkt sammelt. Auf diese Weise

⁴⁵ Anhang VIII: Häufigkeitsverteilungen (Pilotstudie), VIII.a Rohskala „Generalisiertes Vertrauen“

6. Pilotstudie

erreichen mehr Befragte den minimalen Item-Score und die Itemschwierigkeit tendiert gegen 0. Für die Items v2.14, v2.18 und v2.19 konnte dieser Effekt aufgrund der überdurchschnittlichen Streuungen etwas abgefangen werden. Die ermittelte Schwierigkeit ist für diese Items weniger typisch als für die übrigen Items mit unterdurchschnittlichen Streuungen. Für diese letztgenannten Items gilt dagegen, dass sich die Antworten der Befragten enger um den Schwierigkeitswert anordnen. Unter Berücksichtigung der Itemvarianz bei der Berechnung des Schwierigkeitsindex resultiert dann eine niedrigere Itemschwierigkeit. (Bühner (2011), S. 222; Fisseni (1997), S. 45)

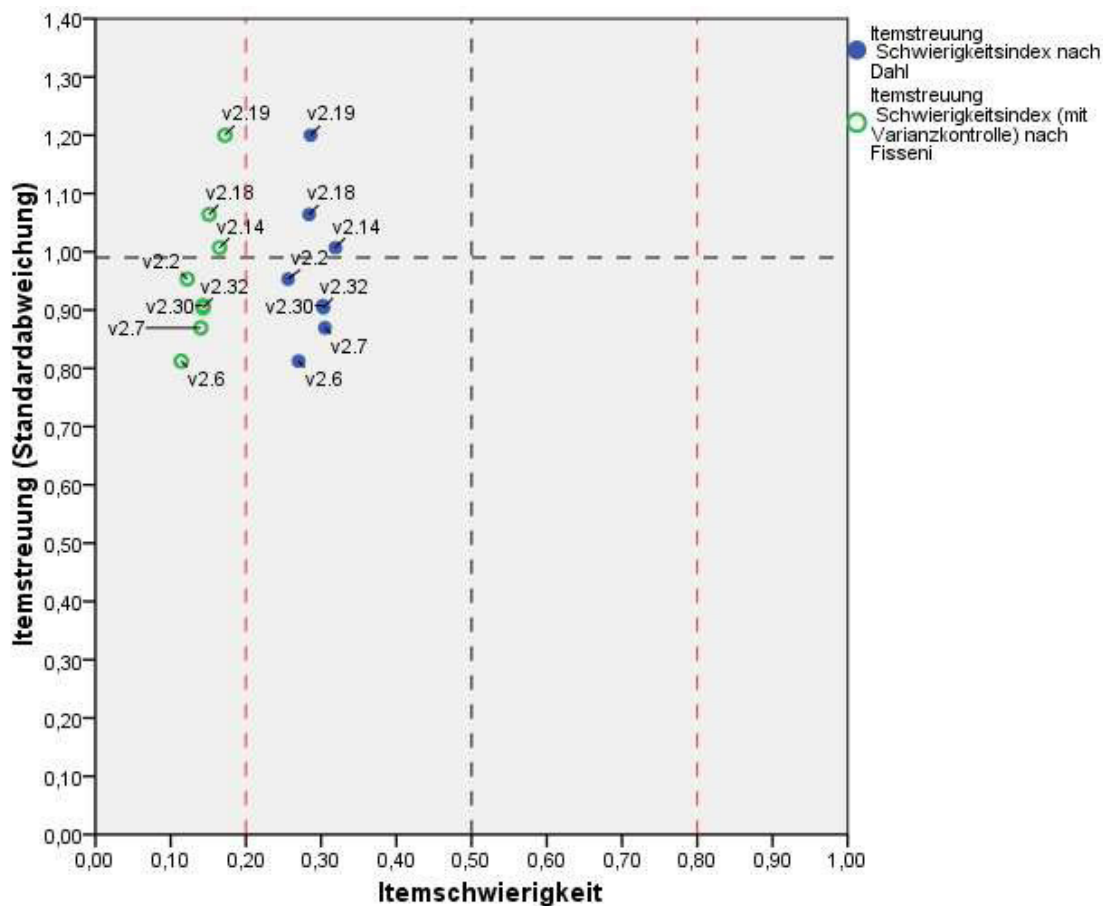


Abb. 6.6: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemstreuung (Standardabweichung) für Items der Rohskala „Generalisiertes Vertrauen“ mit extrem niedriger Schwierigkeit ($P_m < .20$ bzw. $P_{mVar} < .20$) und variabler Streuung; ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) (Bühner (2011), S. 81)

Aus den Histogrammen der selektierten Items lässt sich ablesen, dass sich die zuvor formulierten Annahmen hinsichtlich durchweg linkssteiler Verteilungen bestätigen. Es ergeben sich mit v2.14, v2.18 und v2.19 drei Items, die aufgrund von überdurchschnittlichen Standardabweichungen eher breitipflig verteilt sind, während die übrigen Items unterdurchschnittliche Streuungen aufweisen und dementsprechend schmalipflig verteilt sind. Insgesamt zeigt sich auch auf grafischem Wege, dass die fraglichen Items weniger gut zwischen verschiedenen starken Ausprägungen Generalisierten Vertrauens trennen können. Vielmehr ist

6. Pilotstudie

anzunehmen, dass sie eher dazu neigen, das wahre Ausmaß Generalisierten Vertrauens zu unterschätzen.⁴⁶

6.2.1.1.4 Items mit extrem hoher Schwierigkeit und variabler Streuung

Zuletzt werden diejenigen Items betrachtet, die entweder auf dem *Schwierigkeitsindex nach DAHL* oder dem *Schwierigkeitsindex nach FISSENI* Werte im Bereich für hohe Schwierigkeiten ($P_m > .80$ bzw. $P_{mVar} > .80$) erreichen. Es werden somit Items mit mittleren und extrem niedrigen Schwierigkeiten ausgeschlossen. Gleichzeitig bleiben nur solche Items erhalten, bei denen es den Befragten besonders leichtfiel, in Richtung der Zieldimension zu antworten bzw. dem jeweiligen Statement zuzustimmen. (Bühner (2011), S. 220; Moosbrugger/Kelava (2012), S. 81) Es ergeben sich dann die folgenden Items:

Tab. 6.8: Itemkennwerte der Rohskala: extrem hohe Schwierigkeit; variable Streuung (N = 220)

Item	P_i^a	P_m^b	σ_i^c	Sch_i^d	Ex_i^e	P_{mVar}^f	$r_1(x-i)^g$	Sel^h
v2.22	4,2455	,811	,90343	-1,405	2,111	,709	,143*	0,183
v2.36	4,2182	,805	,88500	-1,160	1,341	,696	,196**	0,247

Sortiert nach P_i

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)

^e Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalipflige Verteilung) (Bühner (2006), S. 88f)

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach LIENERT/RAATZ (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

* $\alpha < .05$ (einseitig), ** $\alpha < .01$ (einseitig)

(Quelle: eigene Berechnungen)

Für die ausgewählten Items ergeben sich hinsichtlich der Einordnung der Schwierigkeiten in einen Wertebereich Diskrepanzen zwischen den beiden Schwierigkeitsindizes. So erreichen beide Items auf dem Schwierigkeitsindex nach DAHL Schwierigkeiten im hohen Bereich ($P_m > .80$), während die Schwierigkeiten nach FISSENI im mittleren Schwierigkeitsbereich angesiedelt sind. Psychometrisch besonders schwer war dabei Item v2.22 mit $P_m = .811$ bzw. $P_{mVar} = .709$. Diesem Item haben im Skalenvergleich die meisten Personen zugestimmt.

Die Abweichung zwischen den Schwierigkeitsindizes ist auch hier wieder auf das Zusammenspiel von Itemstreuung und dem Verhältnis der Randhäufigkeiten zurückzuführen. So erreichen deutlich mehr Befragte den maximalen Item-Score verglichen mit dem minimalen Item-Score, was für eine Tendenz der Itemschwierigkeit in Richtung $P_m = 1$ sorgt. Gleichzeitig scheinen die Standardabweichungen der Items noch ausreichend hoch zu sein, um eine entgegengesetzte Tendenz zu bewirken. Das heißt: Die Antworten der Befragten variieren auf einem skalenspezifisch unterdurchschnittlichen Niveau, unterscheiden sich aber noch relativ stark. Allerdings sind die Items deutlich schmalipflig verteilt, sodass die Linksverschiebung der Items nicht übermäßig groß ausfällt. (Bühner (2011), S. 222; Fisseni (1997), S. 45)

⁴⁶ Anhang V: Häufigkeitsverteilungen (Pilotstudie), V.a Rohskala „Generalisiertes Vertrauen“

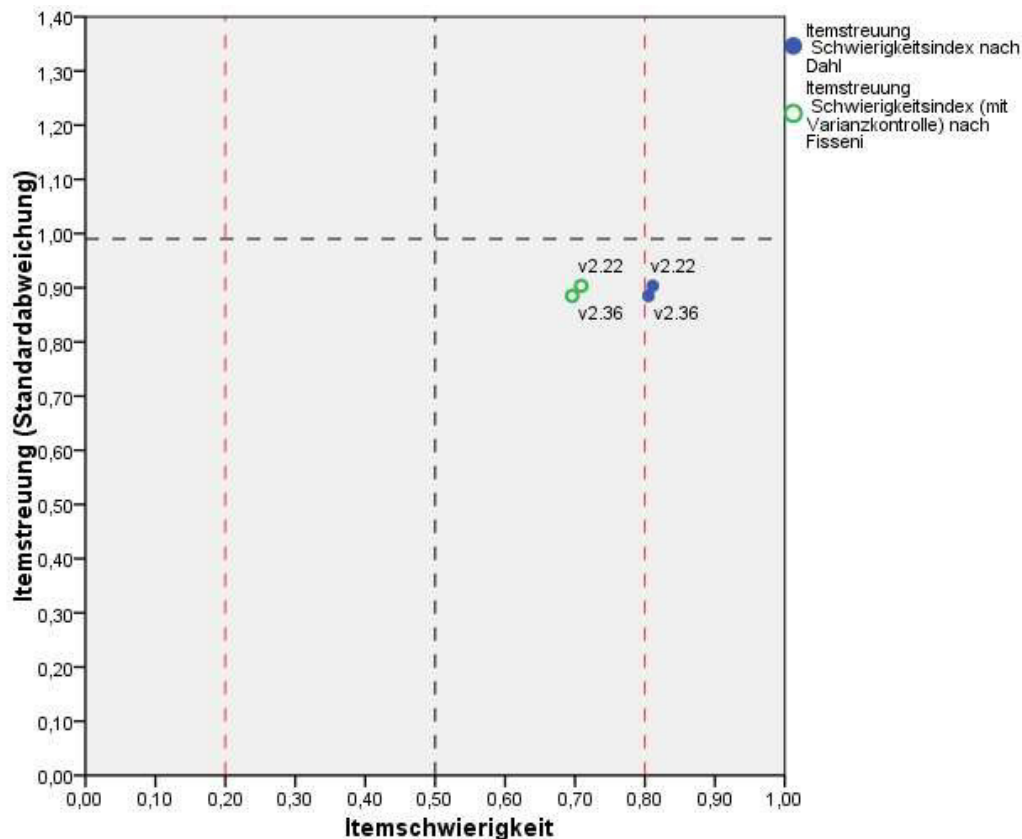


Abb. 6.7: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemstreuung (Standardabweichung) für Items der Rohskala „Generalisiertes Vertrauen“ mit extrem hoher Schwierigkeit ($P_m > .80$ bzw. $P_{mVar} > .80$) und variabler Streuung; ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) (Bühner (2011), S. 81)

Aus den Histogrammen der Items geht hervor, dass sie jeweils deutlich rechtssteil und schmalgipflig verteilt sind. Den Befragten fiel es dementsprechend extrem leicht, auf diesen Items in Schlüsselrichtung der Gesamtskala zu antworten, und sie erreichen eher hohe Vertrauenswerte. Es ist daher zu erwarten, dass diese Items das wahre Ausmaß Generalisierten Vertrauens überschätzen und sie somit für eine differenzierte Messung der Zieldimension ungeeignet sind.⁴⁷

6.2.1.1.1.5 Zusammenfassung der Schwierigkeitsanalyse

Von den 42 Items der vorläufigen Skala „Generalisiertes Vertrauen“ erscheinen nach Abschluss der Schwierigkeitsanalyse insgesamt 14 Items als geeignet, um gut zwischen stark und schwach ausgeprägtem generalisiertem Vertrauen differenzieren zu können. Im Einzelnen handelt es sich dabei um v2.5, v2.10, v2.15, v2.17, v2.23, v2.25 bis v2.27, v2.31, v2.33, v2.35 und v2.40 bis v2.42. Die ausgewählten Items variieren dabei in einem Schwierigkeitsbereich von $P_i = 3,5682$, $P_m = .642$, $P_{mVar} = .490$ bis $P_i = 2,5091$, $P_m = .377$, $P_{mVar} = .210$ und befinden sich damit weitestgehend innerhalb des von FISSENI für homogene Tests vorgeschlagenen Schwierigkeitsintervalls von $P_m = .40-.60$. (Bühner (2006), S. 98ff; Fisseni (1997), S. 60) Gleichzeitig weisen die Items skalenspezifisch überdurchschnittliche Standardabweichungen auf, sodass davon ausgegangen werden kann, dass sich die Antworten der Befragten

⁴⁷ Anhang V: Häufigkeitsverteilungen (Pilotstudie), V.a Rohskala „Generalisiertes Vertrauen“

ausreichend stark voneinander unterscheiden und die Items differenziert messen. (Bühner (2011), S. 220ff; Moosbrugger/Kelava (2012), S. 81ff)

Die Differenzierungsfähigkeit der ausgewählten Items wird auch dadurch sichergestellt, dass sowohl Items mit Schwierigkeiten unterhalb als auch oberhalb des theoretischen Skalenmittelpunktes von 3 einbezogen sind. Zudem ergeben sich für alle selektierten Items breitipflige Häufigkeitsverteilungen. (Fisseni (1997), S. 60)

Die übrigen Items wurden auf der einen Seite ausgeschlossen, weil die jeweiligen Itemschwierigkeiten extrem niedrig oder extrem hoch ($P_m < .20$ bzw. $P_m > .80$) ausfielen und daher anzunehmen war, dass die Items das wahre Ausmaß Generalisierten Vertrauens eher unter- oder überschätzen würden. Auf der anderen Seite wurden auch solche Items ausgeschlossen, die bei mittleren Itemschwierigkeiten unterdurchschnittliche Itemvarianzen aufwiesen. Bezüglich dieser Items war anzunehmen, dass die jeweilige Differenzierungsfähigkeit nicht ausreichend stark ausgeprägt war und die Zieldimension „Generalisiertes Vertrauen“ dementsprechend nicht ausreichend differenziert abgebildet werden würde.

6.2.1.1.2 Trennschärfenanalyse

Die Itemtrennschärfen variieren innerhalb der Rohskala „Generalisiertes Vertrauen“ zwischen $r_{i(x-i)}=.008$ und $r_{i(x-i)}=.465$, wobei 5 Items (ca. 12%) eine Trennschärfe nahe null ($r_{i(x-i)}<.10$) aufweisen. Die Trennschärfen von 20 Items (ca. 48%) sind als niedrig zu bezeichnen ($r_{i(x-i)}<.30$), während die übrigen 17 Items (ca. 40%) Trennschärfen im mittleren Bereich ($r_{i(x-i)}=.30-.50$) erreichen. Keines der Items besitzt eine hohe Itemtrennschärfe ($r_{i(x-i)}>.50$). (Fisseni (1997), S. 124; Bühner (2006), S. 140; Bühner (2011), S. 81, S. 256) Die überwiegende Mehrzahl der Items weist somit eine niedrige Skalenpassung auf. Allerdings werden die Trennschärfen von 34 Items hochsignifikant und von 3 Items signifikant. Unter diesen Items befinden sich sämtliche Items mit mittleren Trennschärfen sowie 20 Items mit niedrigen Trennschärfen, sodass ca. 88% der Items eine mitunter eher schwache, aber trotzdem bedeutsame Item-Skala-Korrelation besitzen. Dies kann allerdings lediglich als ein Hinweis auf eine inhaltliche Homogenität der Skala gewertet werden. (Bühner (2011), S. 246f)

Im Allgemeinen gilt, dass normalerweise mit einer höheren Itemstreuung auch eine höhere Itemtrennschärfe einhergeht bzw. je höher die Standardabweichung eines Items ist, desto höher fällt in der Regel dessen Trennschärfe aus. Eine breitere Verteilung der Antworten der Befragten über die Antwortskala bewirkt somit mitunter eine bessere Passung mit der Gesamtskala bzw. bei großen Unterschieden im Antwortverhalten der Befragten bezüglich eines Items besteht eine höhere Chance, dass die Antworttendenz entlang der Zieldimension der Skala verläuft und sich somit für dieses Item eine höhere Item-Skala-Korrelation ergibt. Hohe Itemvarianzen garantieren allerdings nicht, dass sich auch hohe Trennschärfen einstellen. Der beschriebene Zusammenhang stellt entsprechend keine Gesetzmäßigkeit dar. (Bühner (2011), S. 259; Moosbrugger/Kelava (2012), S. 85) Für die vorläufige Skala „Generalisiertes Vertrauen“ scheint diese Beziehung allerdings weitestgehend keine Gültigkeit zu besitzen.

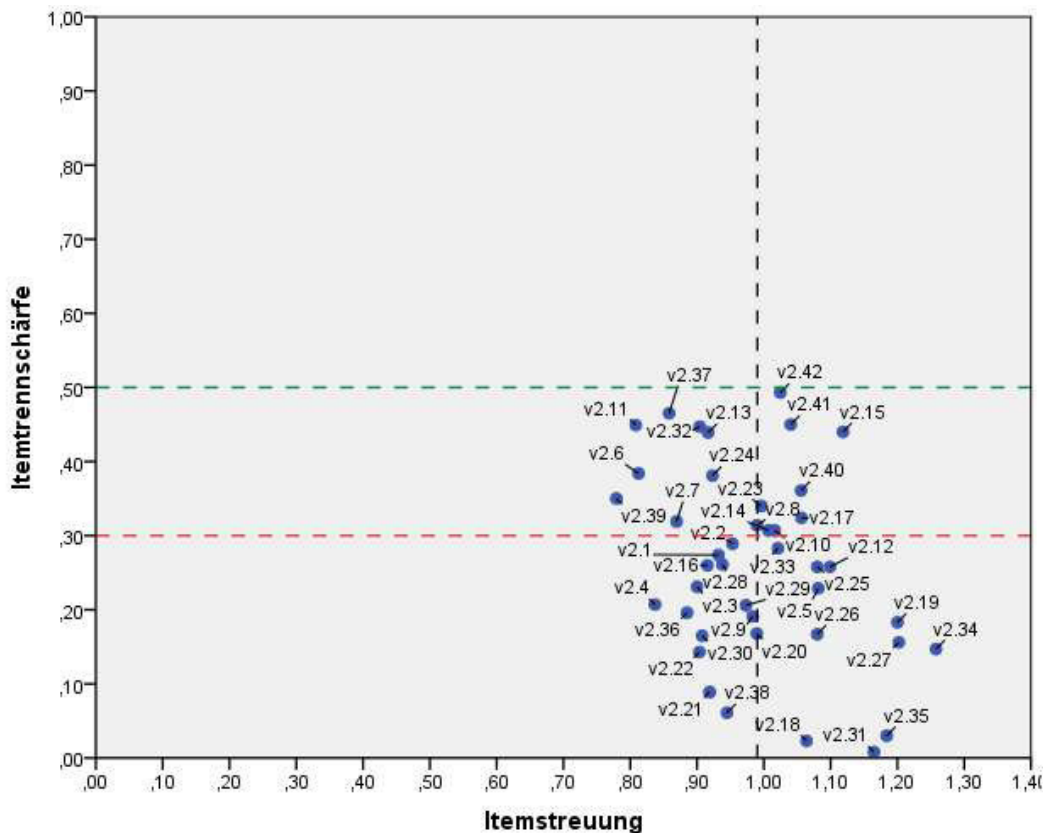


Abb. 6.8: Zusammenhang zwischen Itemstreuung (Standardabweichung) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation [$r_{i(x-i)}$]) für alle Items der Rohskala „Generalisiertes Vertrauen“; ($r_{i(x-i)} < .30$ = Niedrig; $r_{i(x-i)} = .30-.50$ = Mittel; $r_{i(x-i)} > .50$ = Hoch) (Bühner (2011), S. 81)

6.2.1.1.2.1 Items mit extrem niedriger Trennschärfe

Besonders niedrige Trennschärfen lassen sich für die Items v2.18, v2.21, v2.31, v2.35 und v2.38 beobachten. Die genannten Items weisen demnach keine inhaltliche Passung mit der Zieldimension „Generalisiertes Vertrauen“ auf bzw. scheinen ein oder mehrere Konstrukte zu messen, die unabhängig von der inhaltlichen Ausrichtung der Skala sind. Geht man indes doch von einer inhaltlichen Kongruenz mit der Zieldimension aus, so scheinen sich Zustimmung und Ablehnung bei diesen Items die Waage zu halten, sodass sie nur schlecht zwischen hohem und niedrigem „Generalisiertem Vertrauen“ trennen können. Ein Beispiel hierfür könnte Item v2.31 darstellen, das eine Itemschwierigkeit besitzt, die fast deckungsgleich mit dem theoretischen Skalenmittelpunkt ist, eine überdurchschnittliche Streuung aufweist und dabei nahezu symmetrisch verteilt ist. Offensichtlich wird das Item aufgrund dieser Parameter indifferent gegenüber der Zieldimension. (Lienert/Raatz (1998), S. 78; Moosbrugger/Kelava (2012), S. 86;)

Tab. 6.9: Itemkennwerte der Rohskala: extrem niedrige Trennschärfe (N = 220)

Item	P _i ^a	P _m ^b	σ_i^c	Sch _i ^d	Ex _i ^e	P _{mVar} ^f	r _{i (x-y)} ^g	Sel ^h	SK ⁱ
v2.31	3,0636	,516	1,16476	,050	-,781	,351	,008	0,008	0,003
v2.18	2,1364	,284	1,06396	,712	-,175	,151	,023	0,026	0,011
v2.35	3,4045	,601	1,18398	-,345	-,843	,449	,030	0,031	0,013
v2.38	2,8273	,457	,94467	,188	-,716	,264	,061	0,061	0,032
v2.21	4,0227	,756	,91882	-,758	,304	,624	,089	0,104	0,048

Sortiert nach r_{i (x-y)}^a Itemschwierigkeit (Mittelwert)^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)^c Itemstreuung (Standardabweichung)^d Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)^e Exzess bzw. Kurtosis (Ex_i < 0 → breitipflige Verteilung; Ex_i > 0 → schmalipflige Verteilung) (Bühner (2006), S. 88f)^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)^h Selektionskennwert nach LIENERT/RAATZ (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)ⁱ Selektionskennwert nach Bühner (Bühner (2006), S. 99f)^j Durchschnittswert

*α<.05 (einseitig), **α<.01 (einseitig)

(Quelle: eigene Berechnungen)

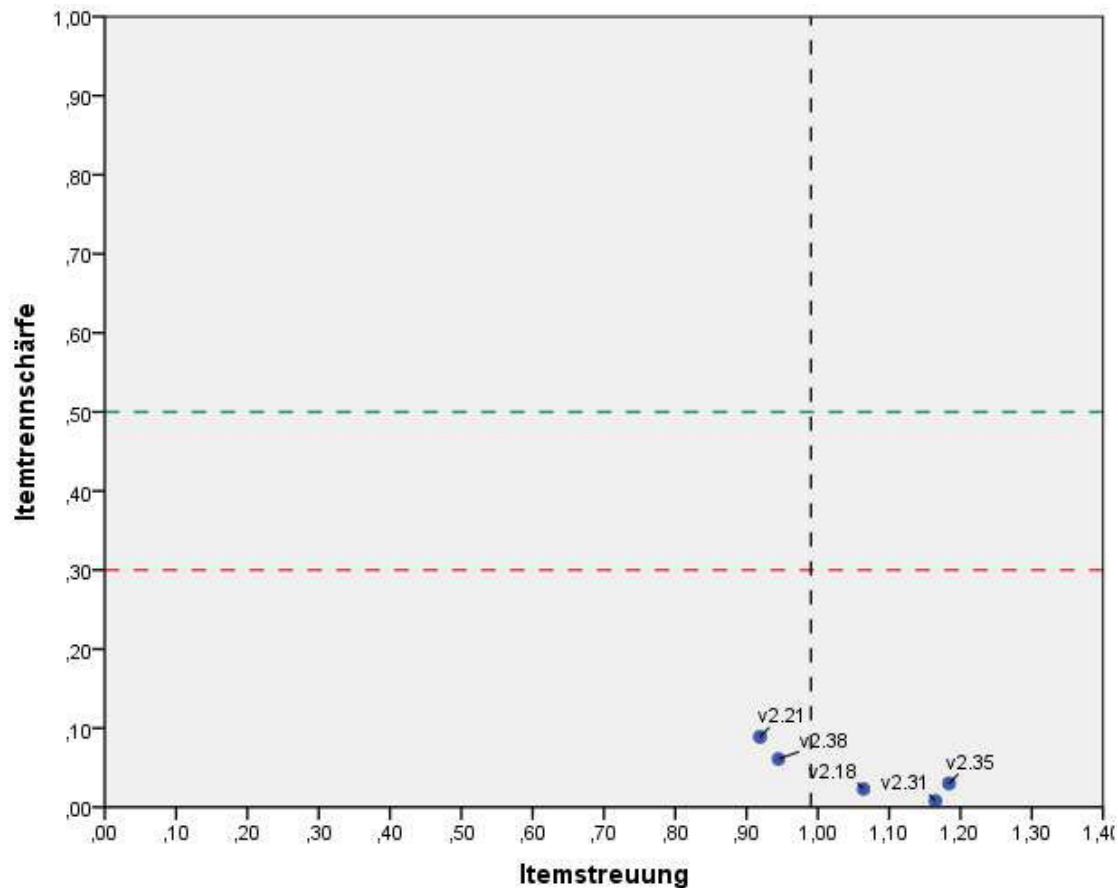


Abb. 6.9: Zusammenhang zwischen Itemstreuung (Standardabweichung) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation [r_{i (x-y)}]) für Items der Rohskala „Generalisiertes Vertrauen“ mit variabler Streuung und extrem niedriger Trennschärfe (r_{i (x-y)}<.10); (r_{i (x-y)}<.30= Niedrig; r_{i (x-y)}=.30-.50= Mittel; r_{i (x-y)}>.50= Hoch) (Bühner (2011), S. 81)

6.2.1.1.2.2 Items mit niedriger, signifikanter Trennschärfe

Niedrige, allerdings signifikante bis hochsignifikante, Trennschärfen finden sich bei den Items v2.1 bis v2.5, v2.9, v2.12, v2.16, v2.19, v2.20, v2.22, v2.25, v2.26 bis v2.30, v2.33, v2.34 und

6. Pilotstudie

v2.36. Die Signifikanzen können dahingehend interpretiert werden, dass die Item-Skala-Korrelation zwar bedeutsam ist, d.h. möglicherweise eine inhaltlich gute Passung vorliegt, aber nur ein geringer Zusammenhang mit der Messrichtung der Skala besteht. Die Antwortmuster der Befragten auf diesen Items entsprechen somit weniger den Antwortmustern, die für die Gesamtskala erwartet werden. Entsprechend ergibt sich eine nur gering ausgeprägte Item-Skala-Passung. (Bühner (2011), S. 247; Lienert/Raatz (1998), S. 78; Moosbrugger/Kelava (2012), S. 86)

Tab. 6.10: Itemkennwerte der Rohskala: niedrige, signifikante Trennschärfe (N = 220)

Item	P ^a	P _m ^b	σ _i ^c	Sch _i ^d	Ex _i ^e	P _{mVar} ^f	r _{i (x-i)} ^g	Sel ^h	SK ⁱ
v2.22	4,2455	,811	,90343	-1,405	2,111	,709	,143*	0,183	0,079
v2.34	3,1273	,532	1,25774	-,159	-1,077	,381	,147*	0,147	0,058
v2.27	3,2818	,570	1,20222	-,288	-,876	,415	,156*	0,158	0,065
v2.30	2,2091	,302	,90756	,757	,552	,143	,165**	0,180	0,091
v2.26	3,3409	,585	1,07998	-,340	-,562	,415	,167**	0,169	0,077
v2.20	3,3273	,582	,98953	-,553	,002	,399	,168**	0,170	0,085
v2.19	2,1455	,286	1,19997	,821	-,342	,172	,183**	0,202	0,076
v2.9	3,2955	,574	,98341	-,622	-,361	,389	,191**	0,193	0,097
v2.36	4,2182	,805	,88500	-1,160	1,341	,696	,196**	0,247	0,111
v2.29	3,8273	,707	,97324	-,667	-,089	,559	,206**	0,226	0,106
v2.4	3,8273	,707	,83703	-,891	1,376	,543	,207**	0,227	0,124
v2.5	2,7227	,431	1,08136	,090	-,691	,258	,229**	0,231	0,106
v2.3	3,2909	,573	,89998	-,267	-,141	,378	,231**	0,233	0,128
v2.12	3,6727	,668	1,09885	-,634	-,465	,522	,258**	0,274	0,117
v2.25	2,8909	,473	1,07987	,021	-,839	,296	,258**	0,258	0,119
v2.16	3,0455	,511	,91548	-,415	-,381	,314	,260**	0,260	0,142
v2.28	2,8182	,455	,93810	-,198	-,547	,261	,261**	0,262	0,139
v2.1	3,7409	,685	,93201	-,212	-,698	,524	,274**	0,295	0,147
v2.33	2,5545	,389	1,02111	,357	-,682	,216	,283**	0,290	0,139
v2.2	2,0227	,256	,95297	,849	,463	,122	,289**	0,331	0,152

Sortiert nach r_{i (x-i)}

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)

^e Exzess bzw. Kurtosis (Ex_i < 0 → breitgipflige Verteilung; Ex_i > 0 → schmalgipflige Verteilung) (Bühner (2006), S. 88f)

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach LIENERT/RAATZ (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

ⁱ Selektionskennwert nach BÜHNER (Bühner (2006), S. 99f)

^j Durchschnittswert

*α<.05 (einseitig), **α<.01 (einseitig)

(Quelle: eigene Berechnungen)

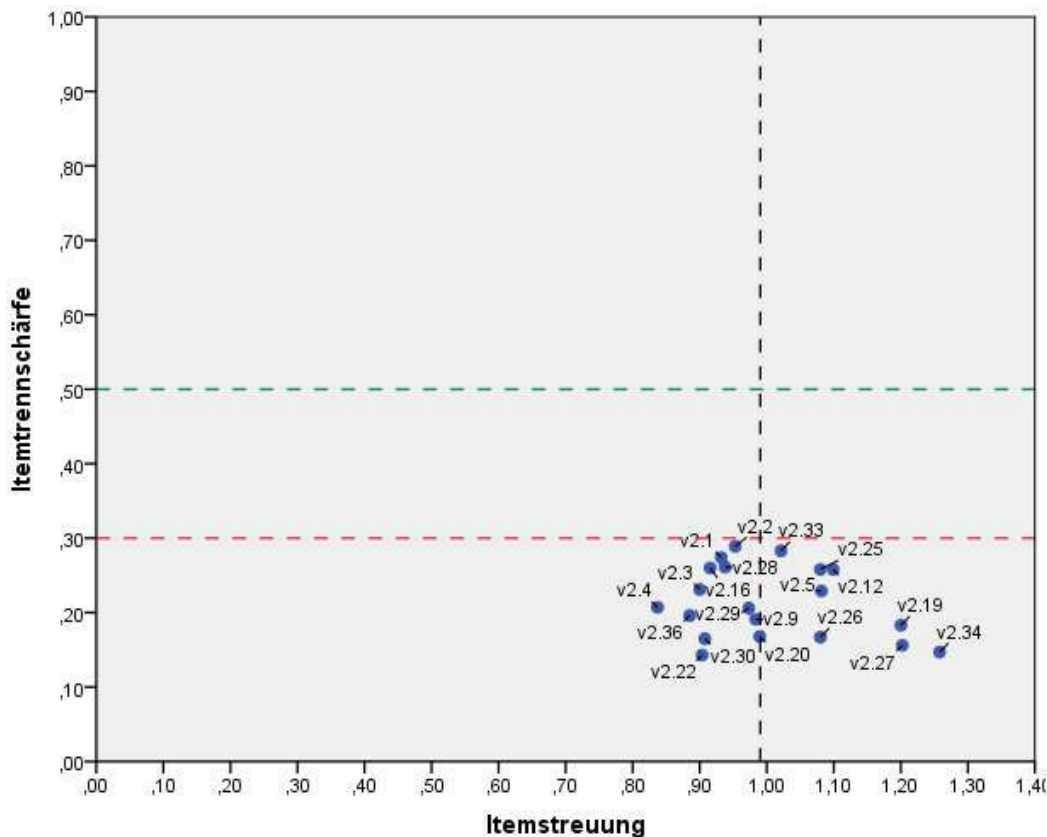


Abb. 6.10: Zusammenhang zwischen Itemstreuung (Standardabweichung) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation $[r_{i(x-i)}]$) für Items der Rohskala „Generalisiertes Vertrauen“ mit variabler Streuung und niedriger, signifikanter Trennschärfe ($r_{i(x-i)} < .30$); ($r_{i(x-i)} < .30$ = Niedrig; $r_{i(x-i)} = .30$ – $.50$ = Mittel; $r_{i(x-i)} > .50$ = Hoch) (Bühner (2011), S. 81)

6.2.1.1.2.3 Items mit mittlerer, hochsignifikanter Trennschärfe

Trennschärfen im unteren mittleren Bereich ($r_{i(x-i)} = .30$ – $.40$) ergeben sich für die Items v2.6 bis v2.8, v2.10, v2.14, v2.17, v2.23, v2.24, v2.39 und v2.40, wobei sich durchweg hochsignifikante Werte einstellen. Die Skalenpassung der Items ist demnach überzufällig und mittelstark ausgeprägt. Betrachtet man die Standardabweichungen, so weisen die Items v2.10, v2.14, v2.17, v2.23 und v2.40 jeweils überdurchschnittliche Varianzen auf, während für die Items v2.6 bis v2.8 sowie v2.24 und v2.39 im Skalenvergleich unterdurchschnittliche Streuungen resultieren. Sowohl für die Items mit vergleichsweise hohen Streuungswerten als auch für diejenigen mit vergleichsweise niedrigen Streuungswerten scheint aber zu gelten, dass das Antwortverhalten der Befragten dem Antwortverhalten für die Gesamtskala ausreichend gut entspricht. Wird auf den einzelnen Items in Richtung der Zieldimension „Generalisiertes Vertrauen“ oder entgegengesetzt dieser Zieldimension geantwortet, so ergibt sich auch für die Gesamtskala tendenziell ein ähnliches Antwortmuster. (Bühner (2011), S. 259)

Dieses Ergebnis ist in der überwiegenden Mehrzahl auch deckungsgleich mit der Polung der genannten Items. So sind die Items v2.8, v2.24, v2.39 und v2.40 positiv, d.h. in Richtung der Zieldimension formuliert, sodass ein hoher Wert ein stark ausgeprägtes Generalisiertes Vertrauen anzeigt. Entsprechend einer positiven Item-Skala-Korrelation ergeben sich dann auch für die Gesamtskala eher hohe Vertrauenswerte. Die Items v2.6, v2.7, v2.10, v2.14 und v2.17 sind dagegen invers zur Zieldimension formuliert. Hier erreichen die Befragten eher niedrige

Werte, was einem eher gering ausgeprägten „Generalisierten Vertrauen“ entspricht. Da diese Items im Vorhinein der Analyse umgepolt wurden, entspricht auch hier eine positive Item-Skala-Korrelation den Erwartungen. Niedrige Werte auf den einzelnen Items lassen niedrige Werte auf der Gesamtskala erwarten. Negativ formulierte Items sind hier demnach gut geeignet, schwaches Generalisiertes Vertrauen zu erfassen, während positiv formulierte Items eher geeignet sind, starkes Generalisiertes Vertrauen zu erfassen. (Moosbrugger/Kelava (2012), S. 86)

Item v2.23 entspricht augenscheinlich zunächst nicht den hier formulierten Modellvorstellungen. So ist das Item zwar negativ formuliert, weist aber ein nahezu symmetrisches Antwortmuster mit einer schwachen Tendenz in positiver Richtung, d.h. in Richtung Generalisierten Vertrauens, auf. Die beschriebene Tendenz ist dabei nur sehr schwach ausgeprägt, allerdings durchaus erkennbar. Somit wurde das Item eher entgegengesetzt der intendierten Richtung, aber in Richtung der Zieldimension „Generalisiertes Vertrauen“ beantwortet. Die Item-Skala-Passung ist demnach ausreichend stark, allerdings scheint dieses negativ gepolte Item verhältnismäßig besser geeignet zu sein, in positiver Skalenrichtung zu messen. (Bühner (2011), S. 176f)

Die Items v2.11, v2.13, v2.15, v2.32, v2.37, v2.41 und v2.42 weisen jeweils hochsignifikante Trennschärfen im oberen mittleren Bereich ($r_{i(x-i)}=.40-.50$) auf. Damit liegen sie zwar noch unterhalb der von FISSENI festgelegten Grenze für hohe Trennschärfen ($r_{i(x-i)}>.50$), aber im von MOOSBRUGGER/KELAVA bestimmten Bereich für „gute“ Trennschärfen ($r_{i(x-i)}=.40-.70$). Die Items besitzen demnach eine bedeutsame und mittelstark bis gute Skalenpassung. (Bühner (2011), S. 81, Fisseni (1997), S. 124, Moosbrugger/Kelava (2012), S. 86) Für die Items v2.15, v2.41 und v2.42 ergeben sich im Skalenvergleich überdurchschnittliche Standardabweichungen, während die Items v2.13, v2.32 und v2.37 unterdurchschnittliche Streuungswerte besitzen. Erstgenannte Items entsprechen somit der Annahme, dass Items mit hohen Standardabweichungen, d.h. einem stark differenzierten Antwortmuster, eine größere Wahrscheinlichkeit besitzen, mit der Gesamtskala zu korrelieren. Für die Items mit unterdurchschnittlichen Varianzen scheint indes zu gelten, dass die Antwortmuster auf den einzelnen Items der für die Gesamtskala zu erwartenden Antwortverteilung weitestgehend entsprechen. (Bühner (2011), S. 259)

Auch für die meisten der hier betrachteten Statements ergibt sich wieder eine Verbindung zwischen Itempolung, Antwortverhalten und Trennschärfe. Die Items v2.11, v2.13 und v2.37 wurden jeweils in Richtung der Zieldimension „Generalisiertes Vertrauen“ gepolt und auch die Antwortmuster auf diesen Items weisen in diese Richtung. In Verbindung mit der positiven Item-Skala-Korrelation sollten hohe Vertrauenswerte auf diesen Items mit hohen Vertrauenswerten auf der Gesamtskala einhergehen. Die Items v2.32, v2.41 und v2.42 wurden dagegen negativ formuliert. Für diese Statements ergaben sich vornehmlich Bewertungen, die entgegengesetzt der Zieldimension gerichtet waren. Da die Items für die Analyse in Richtung der Zieldimension rekodiert wurden, stehen niedrige Werte für entsprechend schwach ausgeprägtes Generalisiertes Vertrauen. Niedrige Vertrauenswerte auf diesen Items sollten, entsprechend der positiven Trennschärfe, gemeinsam mit niedrigen Vertrauenswerten auf der Gesamtskala auftreten. Auch für diese vergleichsweise trennscharfen Items zeigt sich, dass niedrige Ausprägungen Generalisierten Vertrauens besonders gut von negativ gepolten Items erfasst wurden, während positiv gepolte Items hohe Ausprägungen Generalisierten Vertrauens gut erfassen. (Bühner (2011), S. 176f)

Eine weitere Ausnahme von dieser Regel stellt augenscheinlich Item v2.15 dar. Dieses wurde negativ formuliert, von den Befragten aber stark in Richtung der Zieldimension „Generalisiertes

6. Pilotstudie

Vertrauen“ beantwortet. Es fiel den Befragten somit eher schwer, dem Item in seine intendierte Richtung zuzustimmen, wenn gleich es ihnen eher leicht fiel, auf dem Item der Zieldimension der Gesamtskala zuzustimmen. Durch v2.15 werden demnach trotz seiner negativen Polung besser hohe als niedrige Vertrauenswerte erfasst. Die vergleichsweise hohe Item-Skala-Korrelation attestiert ihm, dass es ähnlich differenziert wie der Gesamttest. (Moosbrugger/Kelava (2012), S. 86)

Ausgehend von der Trennschärfenanalyse scheinen die Items v2.11, v2.13, v2.15, v2.32, v2.37, v2.41 und v2.42 besonders gut geeignet zu sein, um Generalisiertes Vertrauen abbilden zu können. Die genannten Items liegen alle im oberen mittleren Wertebereich für Trennschärfen und gelten nach MOOSBRUGGER/KELAVA als „gute“ Trennschärfen. Zusätzlich werden die Item-Skala-Korrelationen durchweg hochsignifikant. Es kann somit von einer bedeutsamen bzw. überzufälligen Skalenpassung ausgegangen werden.

Noch als ausreichend gut geeignet, um Generalisiertes Vertrauen abbilden zu können, erscheinen die Items v2.6 bis v2.8, v2.10, v2.14, v2.17, v2.23, v2.24, v2.39 und v2.40. Hier ergeben sich hochsignifikante Trennschärfen im unteren mittleren Wertebereich. Zwar ist die Skalenpassung aufgrund der niedrigeren Item-Skala-Korrelationen weniger stark, die ermittelten Signifikanzen lassen allerdings darauf schließen, dass die Skalenpassung trotzdem überzufällig und bedeutsam ist.

Tab. 6.11: Itemkennwerte der Rohskala: mittlere, hochsignifikante Trennschärfe (N = 220)

Item	P _i ^a	P _m ^b	σ_i^c	Sch _i ^d	Ex _i ^e	P _{mVar} ^f	r _{i (x-i)} ^g	Sel ^h	SK ⁱ
v2.10	2,9091	,477	1,01626	-,080	-,533	,292	,307**	0,307	0,151
v2.14	2,2773	,319	1,00702	,476	-,604	,165	,307**	0,329	0,152
v2.8	2,9182	,480	,98974	-,177	-,561	,291	,314**	0,314	0,159
v2.7	2,2182	,305	,86939	,529	,133	,140	,319**	0,347	0,183
v2.17	2,5545	,389	1,05628	,125	-1,061	,220	,324**	0,332	0,153
v2.23	3,0545	,514	,99622	-,082	-,459	,326	,340**	0,340	0,171
v2.39	3,4045	,601	,77929	-,613	,501	,399	,350**	0,357	0,225
v2.40	3,2227	,556	1,05572	-,245	-,467	,378	,361**	0,363	0,171
v2.24	3,3955	,599	,92306	-,129	-,615	,412	,381**	0,389	0,206
v2.6	2,0818	,270	,81237	,725	,887	,114	,384**	0,432	0,236
v2.13	3,0682	,517	,91656	-,136	-,368	,320	,439**	0,439	0,239
v2.15	3,5682	,642	1,11850	-,390	-,702	,490	,440**	0,459	0,197
v2.32	2,2136	,303	,90397	,500	-,097	,143	,447**	0,486	0,247
v2.11	3,0636	,516	,80837	-,169	-,256	,307	,449**	0,449	0,278
v2.41	2,5091	,377	1,04024	,283	-,596	,210	,450**	0,464	0,216
v2.37	3,7500	,688	,85808	-,457	-,107	,518	,465**	0,502	0,271
v2.42	2,8636	,466	1,02460	,251	-,434	,282	,493**	0,494	0,241

Sortiert nach r_{i (x-i)}

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe

^e Exzess bzw. Kurtosis

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach LIENERT/RAATZ (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

ⁱ Selektionskennwert nach BÜHNER (Bühner (2006), S. 99f)

* $\alpha < .05$ (einseitig), ** $\alpha < .01$ (einseitig)

(Quelle: eigene Berechnungen)

Es kann außerdem festgestellt werden, dass die ausgewählten trennscharfen Items Generalisiertes Vertrauen sehr differenziert darstellen können. So ergaben sich trennscharfe Items mit besonders hohen wie niedrigen als auch mittleren Itemschwierigkeiten. Die ausgewählten Items tragen somit bereits dem Umstand Rechnung, dass eine Skala in einem breiten Schwierigkeitsbereich differenzieren soll. Auf diese Weise werden alle

6. Pilotstudie

Eigenschaftsbereiche der Befragten, auch die Randbereiche eines Merkmals, trennscharf erfasst. (Bühner (2006), S. 99; Bühner (2011), S. 223, S. 248)

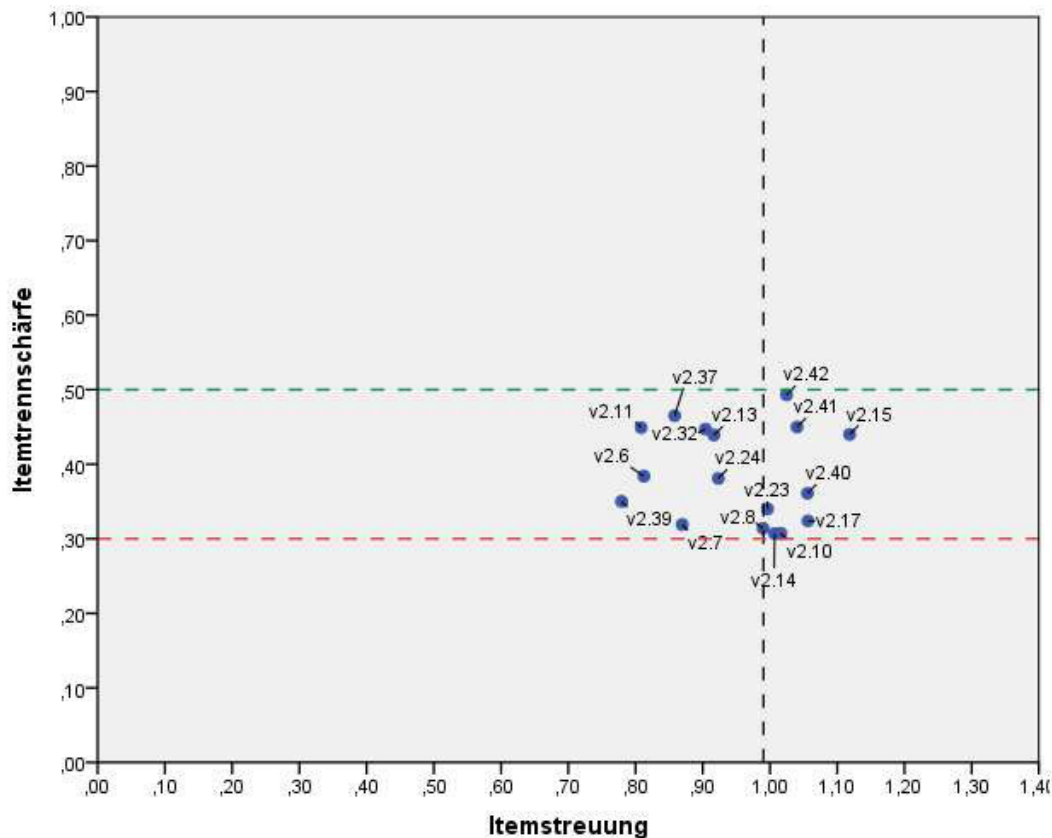


Abb. 6.11: Zusammenhang zwischen Itemstreuung (Standardabweichung) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation $[r_{i(x-i)}]$) für Items der Rohskala „Generalisiertes Vertrauen“ mit variabler Streuung und mittlerer, hochsignifikanter Trennschärfe ($r_{i(x-i)}=.30-.50$); ($r_{i(x-i)}<.30$ = Niedrig; $r_{i(x-i)}=.30-.50$ = Mittel; $r_{i(x-i)}>.50$ = Hoch) (Bühner (2011), S. 81)

Für die übrigen Items wurden lediglich Trennschärfen nahe null bzw. im unteren Wertebereich gemessen. Zwar ergeben sich auch dort teilweise signifikante sowie hochsignifikante Zusammenhänge zwischen den Items und der Gesamtskala, allerdings sind diese zu schwach ausgeprägt, als dass man davon ausgehen könnte, dass diese Items geeignet sind Generalisiertes Vertrauen ausreichend differenziert und gut messen zu können. (Bühner (2011), S. 171f)

6.2.1.1.3 Itemselektion

Ausgehend von den bisherigen Analyseschritten scheinen insgesamt 24 Items der Rohskala „Generalisiertes Vertrauen“ geeignet zu sein, entweder gut zwischen verschiedenen starken Ausprägungen Generalisierten Vertrauens differenzieren zu können, d.h. sie weisen mittlere Itemschwierigkeiten ($P_m = .20-.80$ bzw. $P_{mVar} = .20-.80$) auf, und/oder die Zieldimension „Generalisiertes Vertrauen“ mindestens ausreichend gut abbilden zu können. Das heißt: Sie weisen mindestens Itemtrennschärfen im mittleren Bereich ($r_{i(x-i)} = .30-.50$) auf.

Tab. 6.12: Itemkennwerte der Rohskala: mittlere Schwierigkeit und/oder mittlere Trennschärfe (N = 220)

Item	P _i ^a	P _m ^b	σ_i ^c	Sch _i ^d	Ex _i ^e	P _{mVar} ^f	r _{i (x-y)} ^g	Sel ^h	SK ⁱ
v2.31	3,0636	,516	1,16476	,050	-,781	,351	,008	0,008	0,003
v2.35	3,4045	,601	1,18398	-,345	-,843	,449	,030	0,031	0,013
v2.27	3,2818	,570	1,20222	-,288	-,876	,415	,156*	0,158	0,065
v2.26	3,3409	,585	1,07998	-,340	-,562	,415	,167**	0,169	0,077
v2.5	2,7227	,431	1,08136	,090	-,691	,258	,229**	0,231	0,106
v2.25	2,8909	,473	1,07987	,021	-,839	,296	,258**	0,258	0,119
v2.33	2,5545	,389	1,02111	,357	-,682	,216	,283**	0,290	0,139
v2.10	2,9091	,477	1,01626	-,080	-,533	,292	,307**	0,307	0,151
v2.14	2,2773	,319	1,00702	,476	-,604	,165	,307**	0,329	0,152
v2.8	2,9182	,480	,98974	-,177	-,561	,291	,314**	0,314	0,159
v2.7	2,2182	,305	,86939	,529	,133	,140	,319**	0,347	0,183
v2.17	2,5545	,389	1,05628	,125	-1,061	,220	,324**	0,332	0,153
v2.23	3,0545	,514	,99622	-,082	-,459	,326	,340**	0,340	0,171
v2.39	3,4045	,601	,77929	-,613	,501	,399	,350**	0,357	0,225
v2.40	3,2227	,556	1,05572	-,245	-,467	,378	,361**	0,363	0,171
v2.24	3,3955	,599	,92306	-,129	-,615	,412	,381**	0,389	0,206
v2.6	2,0818	,270	,81237	,725	,887	,114	,384**	0,432	0,236
v2.13	3,0682	,517	,91656	-,136	-,368	,320	,439**	0,439	0,239
v2.15	3,5682	,642	1,11850	-,390	-,702	,490	,440**	0,459	0,197
v2.32	2,2136	,303	,90397	,500	-,097	,143	,447**	0,486	0,247
v2.11	3,0636	,516	,80837	-,169	-,256	,307	,449**	0,449	0,278
v2.41	2,5091	,377	1,04024	,283	-,596	,210	,450**	0,464	0,216
v2.37	3,7500	,688	,85808	-,457	-,107	,518	,465**	0,502	0,271
v2.42	2,8636	,466	1,02460	,251	-,434	,282	,493**	0,494	0,241

Sortiert nach r_{i (x-y)}

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)

^e Exzess bzw. Kurtosis (Ex_i < 0 → breitgipflige Verteilung; Ex_i > 0 → schmalgipflige Verteilung) (Bühner (2006), S. 88f)

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach LIENERT/RAATZ (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

ⁱ Selektionskennwert nach BÜHNER (Bühner (2006), S. 99f)

*p<.05 (einseitig), **p<.01 (einseitig)

(Quelle: eigene Berechnungen)

Es zeigt sich, dass nicht alle der aus Schwierigkeits- und Trennschärfenanalyse resultierenden Items gleichzeitig ausreichend trennscharf sind, ausreichend gut zwischen hohen und niedrigen Werten trennen können und eine ausreichend ausgeprägte Differenzierungsfähigkeit besitzen. So ergeben sich für die Items v2.6, v2.7, v2.14 und v2.32 zwar Trennschärfen im mittleren Bereich, allerdings weisen sie auf dem *Schwierigkeitsindex von FISSENI* nur niedrige Schwierigkeiten auf. Dagegen weisen die Items v2.5, v2.25 bis v2.27, v2.31, v2.33 und v2.35 jeweils günstige Schwierigkeiten im mittleren Bereich auf, erreichen aber nur niedrige Trennschärfen.

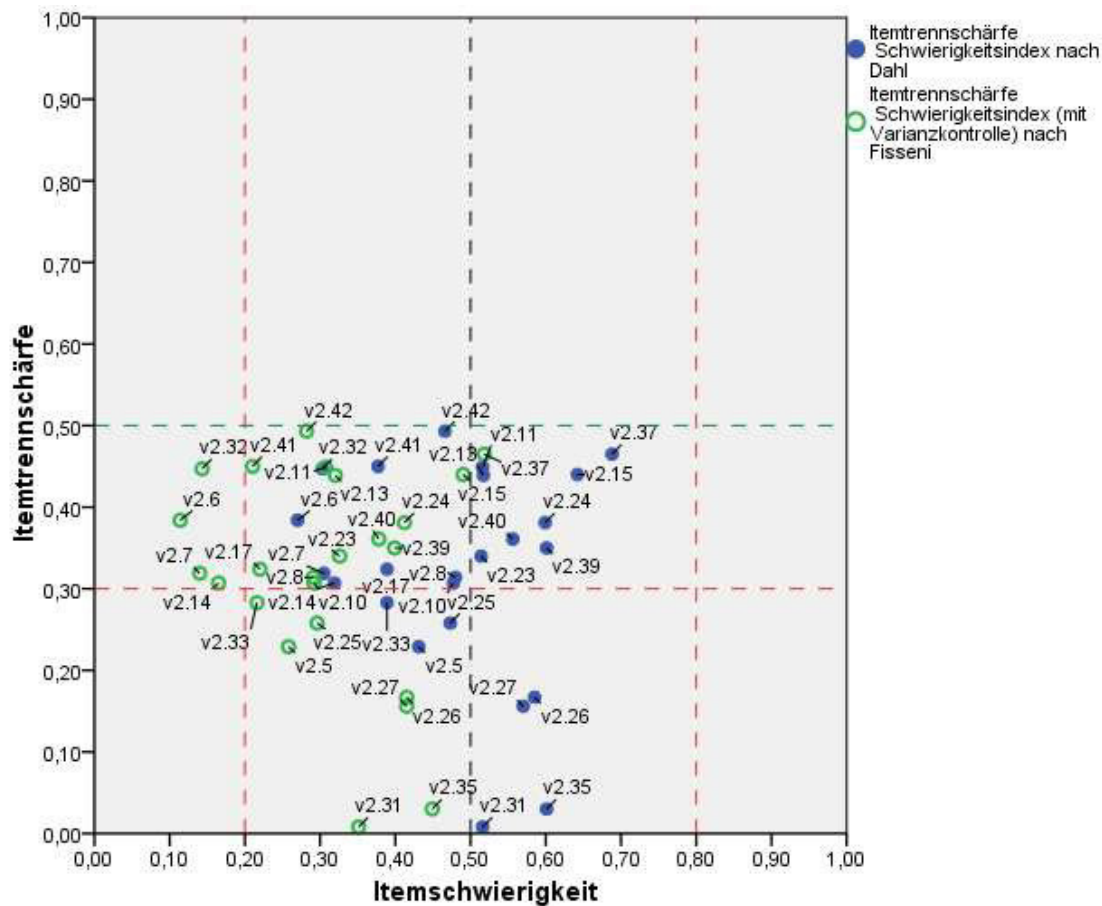


Abb. 6.12: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation [$r_{i(x-i)}$]) für Items der Rohskala „Generalisiertes Vertrauen“ mit mittlerer Schwierigkeit ($P_m = .20-.80$ bzw. $P_{mVar} = .20-.80$) auf mindestens einem der Indizes von DAHL oder FISSENI und/oder mittlerer, hochsignifikanter Trennschärfe ($r_{i(x-i)} = .30-.50$); ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) und ($r_{i(x-i)} < .30$ = Niedrig; $r_{i(x-i)} = .30-.50$ = Mittel; $r_{i(x-i)} > .50$ = Hoch) (Bühner (2011), S. 81)

Wie bereits beschrieben, soll aus diesen Items ein möglichst homogener Test generiert werden. Ziel ist es somit, solche Items zu selektieren, die nach Möglichkeit dasselbe Merkmal erfassen. Hierzu sollten sich die Itemschwierigkeiten in einem relativ engen Bereich bewegen bzw. nur geringe Unterschiede aufweisen. Da Items mit mittleren Schwierigkeiten, die eng um den theoretischen Skalenmittelpunkt streuen, in der Regel die höchsten Standardabweichungen aufweisen und damit eine besonders stark ausgeprägte Differenzierungsfähigkeit besitzen, bietet sich hier ein Schwierigkeitsintervall mit $P_m = .40-.60$ an. Des Weiteren sollten vornehmlich solche Items ausgewählt werden, die möglichst maximale Trennschärfen aufweisen. Hohe Trennschärfen zeigen nicht nur einen starken Zusammenhang mit der Zieldimension der Skala an, sondern begünstigen auch hohe Iteminterkorrelationen. Inhaltlich bedeuten diese, dass die Items nicht unbedingt identische, aber sehr ähnliche Aspekte eines Merkmals erfassen. (Bühner (2006), S. 98ff; Bühner (2011), S. 248f, S. 258; Fisseni (1997), S. 54ff, S. 60ff; Lienert/Raatz (1998), S. 36, S. 100, S. 115; Moosbrugger/Kelava (2012), S. 32f, S. 81ff)

Primäres Auswahlkriterium bei der Erstellung eines homogenen Tests ist die Itemtrennschärfe. Die Itemschwierigkeit stellt dagegen lediglich ein sekundäres Kriterium dar. So werden nachfolgend Items mit niedrigen Trennschärfen ausgeschlossen, auch wenn sie mittlere bzw. günstige Schwierigkeiten aufweisen. Um allerdings auch zwischen extremen Merkmalsausprägungen trennen zu können, sollen auch Items mit vergleichsweise niedrigen

6. Pilotstudie

oder vergleichsweise hohen Itemschwierigkeiten beibehalten werden, wenn sie ausreichend hohe Trennschärfen besitzen. Es bleiben somit auch solche Items erhalten, die bei einer hohen Trennschärfe vergleichsweise extreme Schwierigkeiten aufweisen. (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 115)

Wendet man die obengenannte Auswahlstrategie auf die Items aus den vorhergehenden Analyseschritten an, so werden sämtliche Items mit niedrigen Trennschärfen ($r_{i(x-i)} < .30$) ausgeschlossen. Beibehalten werden somit insgesamt 17 Items, die teilweise auch extreme Itemschwierigkeiten besitzen. Dabei weisen die Items v2.6, v2.7, v2.14 und v2.32 jeweils nur auf dem Schwierigkeitsindex nach DAHL mittlere Itemschwierigkeiten auf. Außerdem ergeben sich für die Items v2.6 bis v2.8, v2.11, v2.12, v2.24, v2.32, v2.37 und v2.39 nur unterdurchschnittliche Standardabweichungen, was eine zumindest eingeschränkte Differenzierungsfähigkeit anzeigt.

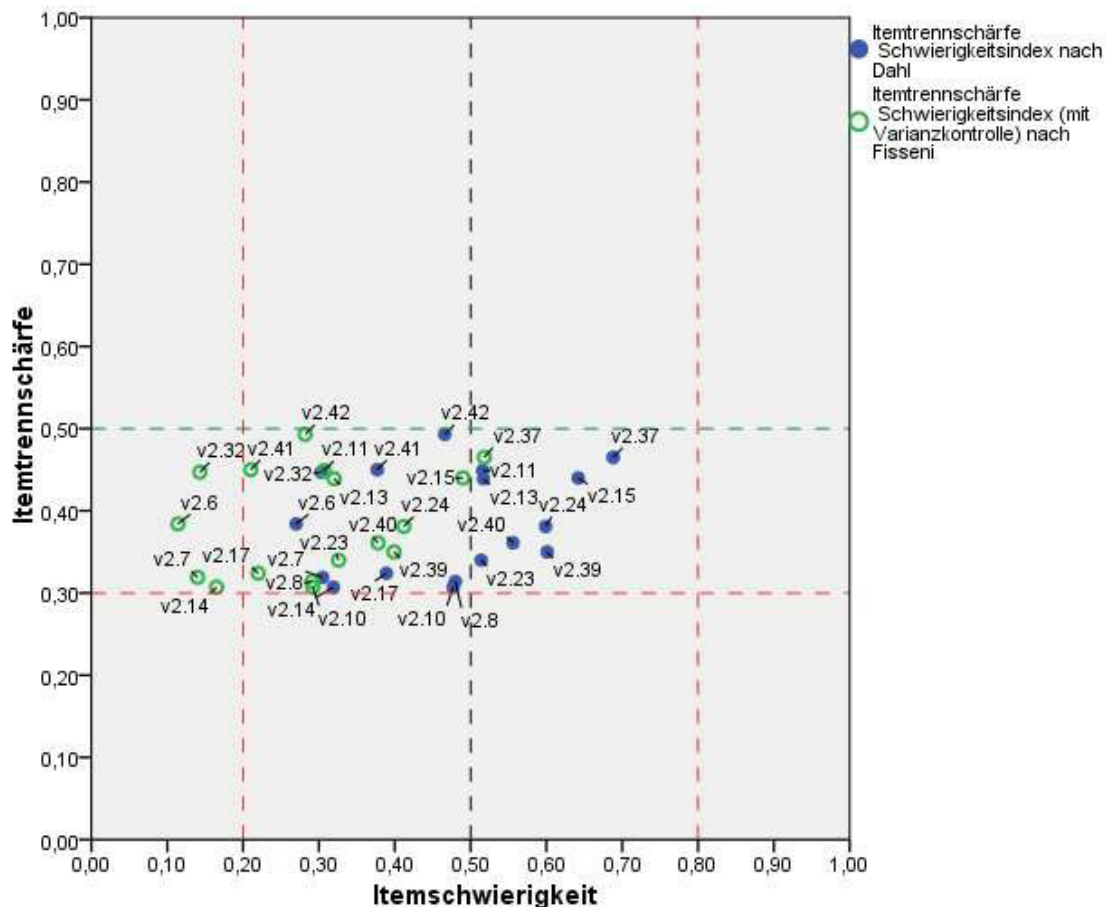


Abb. 6.13: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation [$r_{i(x-i)}$]) für Items der Rohskala „Generalisiertes Vertrauen“ mit mittlerer Schwierigkeit ($P_m = .20-.80$ bzw. $P_{mVar} = .20-.80$) auf mindestens einem der Indizes von DAHL oder FISSENI und mittlerer, hochsignifikanter Trennschärfe ($r_{i(x-i)} = .30-.50$); ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) und ($r_{i(x-i)} < .30$ = Niedrig; $r_{i(x-i)} = .30-.50$ = Mittel; $r_{i(x-i)} > .50$ = Hoch) (Bühner (2011), S. 81)

Lediglich die Items v2.10, v2.15, v2.17, v2.23 und v2.40 bis v2.42 erfüllen sowohl die Vorgaben hinsichtlich mittlerer Itemschwierigkeiten mit ausreichend ausgeprägter Differenzierungsfähigkeit und Trennschärfen, die mindestens im mittleren Bereich liegen. Diese Items sollten somit besonders gut geeignet sein, um Generalisiertes Vertrauen trennscharf und differenziert messen zu können.

6. Pilotstudie

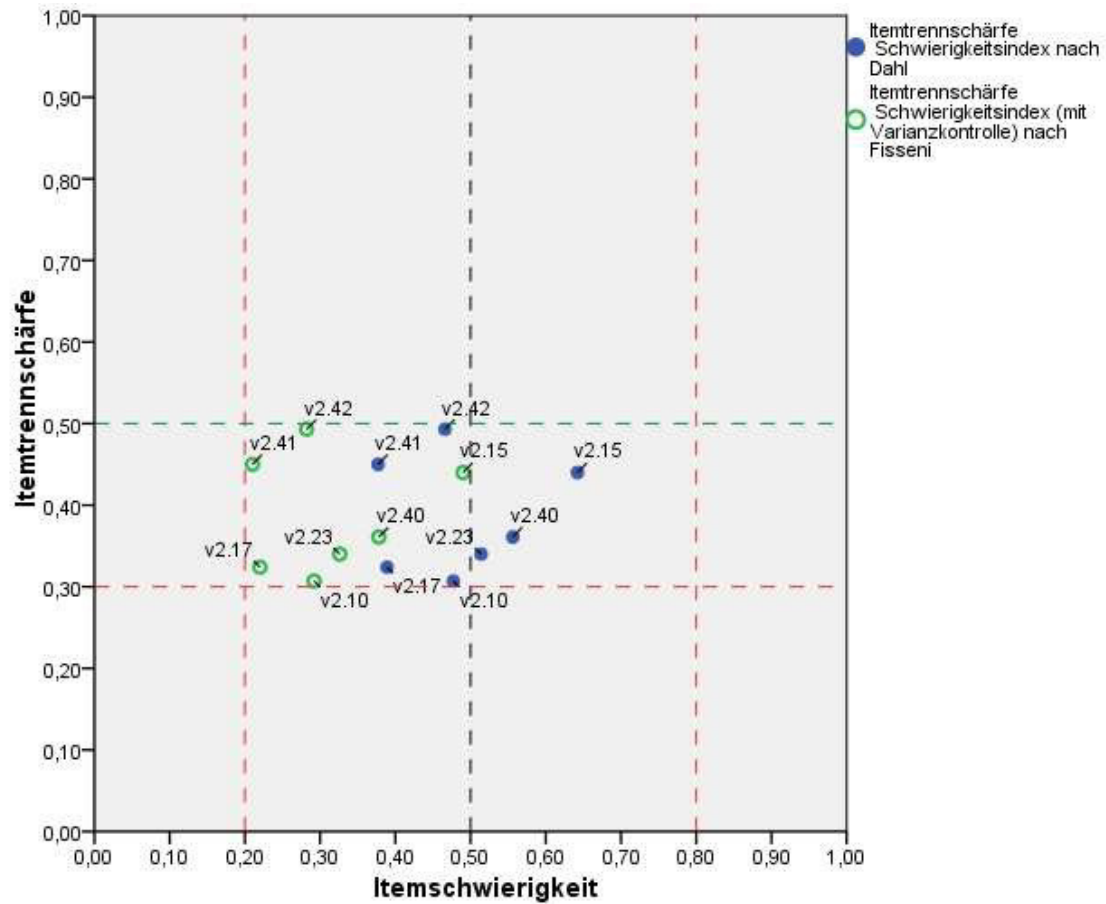


Abb. 6.14: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation [$r_{i(x-i)}$]) für Items der Rohskala „Generalisiertes Vertrauen“ mit mittlerer Schwierigkeit ($P_m = .20-.80$ bzw. $P_{mVar} = .20-.80$) auf den Indizes von DAHL oder FISSENI und mittlerer, hochsignifikanter Trennschärfe ($r_{i(x-i)} = .30-.50$); ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) und ($r_{i(x-i)} < .30$ = Niedrig; $r_{i(x-i)} = .30-.50$ = Mittel; $r_{i(x-i)} > .50$ = Hoch) (Bühner (2011), S. 81)

Besonders hervorzuheben ist dabei Item v2.42, welches als einziges Item sowohl die strengen Kriterien von FISSENI ($P_m = .40-.60$) als auch von MOOSBRUGGER/KELAVA ($r_{i(x-i)} = .40-.70$) erfüllt. (Fisseni (1997), S. 60; Moosbrugger/Kelava (2012), S. 86)

6. Pilotstudie

Aus der Itemanalyse resultiert demnach die sogenannte Objektiv-Skala, die Generalisiertes Vertrauen abbildet und sich aus den folgenden Items zusammensetzt:

Tab. 6.13: Resultat der Itemanalyse für Objektiv-Skala

(1)	2.6) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen. (-)
(2)	2.7) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken. (-)
(3)	2.8) Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden. (+)
(4)	2.10) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinandertreffen, ist von andauernden Konflikten geprägt. (-)
(5)	2.11) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden. (+)
(6)	2.13) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen. (+)
(7)	2.14) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen. (-)
(8)	2.15) Am sichersten fährt man mit der Annahme, dass alle Menschen einen bösartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt. (-)
(9)	2.17) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte. (-)
(10)	2.23) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden. (-)
(11)	2.24) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist. (+)
(12)	2.32) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun. (-)
(13)	2.37) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden. (+)
(14)	2.39) Wenn man einen Gegenstand an eine andere Person verleiht, wird man diesen intakt zurückbekommen. (+)
(15)	2.40) Heutzutage ist es nicht schwer Freunde zu finden. (+)
(16)	2.41) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt. (-)
(17)	2.42) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern. (-)

Hoher Itemscore entspricht starker Ausprägung der Zieldimension, niedriger Itemscore entspricht schwacher Ausprägung der Zieldimension. Es ergeben sich daher die folgenden Codierungen für die Antwortskala:

(+) Itempolung positiv: „Stimme nicht zu“ (0); „Stimme eher zu“ (1); „Stimme teils zu, stimme teils nicht zu“ (2); „Stimme eher zu“ (3); „Stimme zu“ (4)

(-) Itempolung negativ: „Stimme nicht zu“ (4); „Stimme eher zu“ (3); „Stimme teils zu, stimme teils nicht zu“ (2); „Stimme eher zu“ (1); „Stimme zu“ (0)

6.2.2 Reliabilitätsanalyse (Objektiv-Skala)

Rechnerisch ist die Reliabilität als Varianzverhältnis zwischen den messfehlerfreien Werten eines Tests und den empirisch ermittelten Werten eines Tests definiert. Eine Schätzung dieses Varianzverhältnisses erfolgt anhand von Korrelationen zwischen Mehrfachmessungen mit dem fraglichen Untersuchungsinstrument. Dabei existieren verschiedene Verfahren zur Reliabilitätsschätzung: Retestmethode, Paralleltestmethode und Konsistenzmethode. (Bühner (2011), S. 153ff; Moosbrugger/Kelava (2012), S. 121f)

Zur Bestimmung der Retest-Reliabilität wird ein Test derselben Stichprobe mit einem gewissen Zeitabstand wiederholt vorgelegt. Es gilt dann:

„Die Retest-Reliabilität eines Testverfahrens ist hoch, wenn zwei Messungen mit diesem Test zu verschiedenen Messzeitpunkten hoch miteinander korrelieren.“ (Moosbrugger/Kelava (2012), S. 123)

Der besondere Vorteil dieser Methode besteht darin, dass sie Aussagen über die Merkmalsstabilität machen kann. Der Abstand der Messzeitpunkte kann dabei allerdings einen erheblichen Einfluss auf die Höhe der ermittelten Reliabilität haben. So können beispielsweise Verzerrungen durch zwischenzeitliches Geschehen oder Erinnerungs- und Übungseffekte entstehen. Da für die hier zu prüfende Skala aufgrund organisatorischer Beschränkungen und des zu erwartenden unverhältnismäßigen Aufwandes keine zweite Messung durchgeführt

werden konnte, kommt diese Form der Reliabilitätsschätzung nachfolgend nicht zum Einsatz. (Bühner (2011), S. 60f, S. 159; Moosbrugger/Kelava (2012), S. 122ff)

Für die Schätzung der Paralleltest-Reliabilität werden zunächst zwei sogenannte Parallelformen des zu analysierenden Tests benötigt. Parallelformen stellen dabei „voneinander verschiedene Varianten des gleichen Testverfahrens, die das gleiche Merkmal mit der gleichen Genauigkeit erfassen“ (Moosbrugger/Kelava (2012), S. 125f) dar. Die beiden Tests werden dann derselben Stichprobe zur Bearbeitung vorgelegt. (Bühner (2006), S. 129; Bühner (2011), S. 158) Es gilt dann:

„Die Paralleltest-Reliabilität eines Testverfahrens ist hoch, wenn zwei parallele Testformen dieses Verfahrens (die gleiche wahre Werte und Fehlervarianzen aufweisen) hoch miteinander korrelieren.“ (Moosbrugger/Kelava (2012), S. 126)

Die Verwendung dieser Methode zu Reliabilitätsbestimmung ermöglicht es zu prüfen, ob die Testergebnisse vor dem Hintergrund einer Variation der Itemauswahl stabil bleiben. Es wird die sogenannte Bedingungsstabilität untersucht. Vorteil dieser Methode ist ihre Unempfindlichkeit gegenüber Erinnerungseffekten. Allerdings können Verletzungen perfekter Parallelität der Testteile Fehleinschätzungen der Reliabilität bewirken. Oft stehen für die Erstellung exakt paralleler Testformen kaum genug gleichermaßen gut geeignete Items zur Verfügung. In Verbindung mit dem kaum gerechtfertigten Aufwand für die Konstruktion zweier zueinander äquivalenter Tests, wurde deshalb darauf verzichtet, Parallelformen für die hier zu prüfende Skala zu entwickeln. Entsprechend wird auch die Paralleltestmethode an dieser Stelle nicht zur Schätzung der Reliabilität eingesetzt werden. (Bühner (2011), S. 61, S. 158; Moosbrugger/Kelava (2012), S. 126ff)

Die Konsistenzmethode unterteilt sich in zwei Verfahren und stellt die am häufigsten für Reliabilitätsschätzungen herangezogene Technik dar. Außerdem sind die Verfahren der Konsistenzanalyse gute Schätzer der Reliabilität, während die wahre Reliabilität durch die Retest- bzw. die Paralleltestmethode auch unter- oder überschätzt wird. Zusätzlich sind Konsistenzanalysen besonders gut geeignet, wenn durch einen Test homogene Merkmalsbereiche erfasst werden sollen oder nur eine einmalige Messung möglich ist. Die an dieser Stelle zu prüfende Skala wurde den Befragten nur ein einziges Mal vorgelegt und soll außerdem nur ein einziges Merkmal erfassen. Ziel ist es somit, eine homogene und eindimensionale Skala zu erstellen, mit deren Hilfe zuverlässig Generalisiertes Vertrauen gemessen werden kann. (Bühner (2006), S. 35, S. 127ff; Bühner (2011), S. 60, S. 157ff; Moosbrugger/Kelava (2012), S. 130ff)

Gemeinsames Merkmal der verschiedenen Verfahren der Konsistenzanalyse ist, dass sie den zu prüfenden Test in Untertests zerlegen. Im Falle der Split-Half- bzw. Testhalbierungsreliabilität wird der Test in zwei Hälften aufgeteilt und diese jeweils „halben“ Tests miteinander korreliert. Die beiden Testhälften sollten dabei möglichst parallel sein. Hierzu können verschiedene Aufteilungsverfahren angewendet werden (Bühner (2006), S. 127; Bühner (2011), S. 157; Moosbrugger/Kelava (2012), S. 129f):

- (1) *Odd-Even-Methode*: Abwechselnde Zuteilung der Items zu den beiden Testhälften, sodass ein Testteil alle geradzahlig Items und der andere alle ungeradzahlig Items enthält.
- (2) *Zeitpartitionierung bzw. Zeitfraktionierung*: Der Gesamttest wird in zwei Teile mit gleich langer Bearbeitungszeit unterteilt.

(3) *Itemzwillinge*: Die Testteile enthalten jeweils Items mit gleicher oder ähnlicher Schwierigkeit und Trennschärfe.

(4) *Zufällige Aufteilung*: Die Items werden den Testhälften zufällig zugewiesen.

Für alle diese Auswahlmethoden gilt allerdings, dass eine tatsächliche Parallelität oder auch nur Ähnlichkeit der Testhälften nicht sichergestellt werden kann.

Bei der Bestimmung der inneren bzw. internen Konsistenz wird der Test dagegen in so viele Untertests wie Items zerlegt. Es wird in diesem Sinne jedes Item als eigenständige Messung der Zieldimension interpretiert. Es gilt dabei:

„Die interne Konsistenz eines Tests ist umso höher, je höher die Korrelationen zwischen den Items im Durchschnitt sind.“ (Moosbrugger/Kelava (2012), S. 131)

Nachfolgend wird zunächst die innere Konsistenz der Objektiv-Skala geprüft. Dabei werden solche Items entfernt, die den Erwartungen für eine homogene und eindimensionale Skala entgegenstehen. Hierbei ist allerdings einschränkend festzuhalten, dass sich aus der Bestimmung der internen Konsistenz eines Tests allenfalls ein Hinweis auf Homogenität ergibt.

„Internal consistency refers to the interrelatedness of a set of items, whereas *homogeneity* refers to the unidimensionality of the set of items. Internal consistency is certainly necessary for homogeneity, but it is not sufficient.“ (Schmitt (1996), S. 350)

Hieran anschließend wird die Split-Half-Reliabilität der resultierenden Skala bestimmt, um die Ergebnisse hinsichtlich der inneren Konsistenz abzusichern. Die Schätzung der Reliabilität anhand der Testhalbierungsmethode erfolgt erst als zweiter Analyseschritt, da durch die Art und Weise der Aufteilung der Items auf die Testhälften unterschiedliche Reliabilitätsschätzungen resultieren können. (Bühner (2011), S. 166; Cronbach (1951), S. 298)

6.2.2.1 Innere Konsistenz

Die Schätzung der inneren Konsistenz einer Skala erfolgt standardmäßig anhand von *Cronbach- α* , welches die Höhe der mittleren Itemzusammenhänge in Abhängigkeit von der Itemanzahl angibt. Dabei fällt *Cronbach- α* dann besonders hoch aus, wenn die Itemanzahl zunimmt und die Items stark positiv korrelieren. Im Allgemeinen kann *Cronbach- α* Werte zwischen 0 (keine Reliabilität) und 1 (perfekte Reliabilität) annehmen. (Bühner (2006), S. 131ff; Bühner (2011), S. 166ff; Cronbach (1951), S. 331; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 247; Moosbrugger/Kelava (2012), S. 130ff; Schnell/Hill/Esser (2008), S. 153) Darüber, in welchem Wertebereich *Cronbach- α* sich bewegen muss, um eine ausreichende Reliabilität anzuzeigen, herrscht in der Literatur eine gewisse Uneinigkeit. Weitestgehend anerkannt ist allerdings eine untere Grenze von $\alpha = .70$. (Bühner (2011), S. 80; Cortina (1993), S. 101; Schmitt (1996), S. 351; Spector (1992), S. 32) So gelten Werte von $\alpha = .70$ -.80 je nach Kontext als akzeptabel, während von anderen Theoretikern Werte mit $\alpha < .80$ als niedrig angesehen werden. (Bühner (2011), S. 241; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 247; Mummendey/Grau (2014), S. 102) Allgemeiner Konsens scheint dagegen darüber zu bestehen, dass Reliabilitäten von $\alpha > .80$ als gut und $\alpha > .90$ als sehr gut bewertet werden. (Bühner (2011), S. 81; Fisseni (1997), S. 124; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 247; Mummendey/Grau (2014), S. 102)

Nachfolgend soll die Reliabilität der Objektiv-Skala durch eine sogenannte Alpha-Maximierung sukzessive gesteigert werden. Hierzu werden schrittweise Items mit niedrigen Trennschärfen ($r_{i(x-i)} < .30$) aus der Skala entfernt, was eine Erhöhung des *Cronbach- α* -Koeffizienten zur Folge hat. Denn je niedriger die Trennschärfe eines Items, desto weniger gut bildet dieses die Zieldimension der Skala ab und desto weniger gut passt es zu den restlichen Items der Skala. Dieser Prozess wird solange fortgesetzt, bis *Cronbach- α* nicht weiter ansteigt. (Bühner (2006), S. 138; Bühner (2011), S. 249f)

Daneben soll die resultierende Skala aber noch weitere Parameter erfüllen. So soll die resultierende Skala nicht nur eine hohe Reliabilität aufweisen, sondern gleichzeitig auch noch möglichst homogen und eindimensional sein. Dabei zeigt eine hohe Reliabilität noch keine eindimensionale Skala an. Vielmehr kann die innere Konsistenz auch bei einer mehrdimensionalen Skala hoch ausfallen. Verantwortlich hierfür sind vornehmlich die Anzahl der in der jeweiligen Skala zusammengefassten Items sowie die Stärke der Korrelationen zwischen den Items und gegebenenfalls den Dimensionen der Skala. Um einen Hinweis auf die Homogenität und Eindimensionalität der Skala zu erhalten, sollte daher nicht *Cronbach- α* betrachtet werden, sondern es sollten die mittlere Inter-Item-Korrelation (MIC) und die Präzision von α (P_α) als Homogenitätsmaße interpretiert werden.

Nach BRIGGS/CHEEK ist ein optimales Homogenitätslevel bei MIC = .20-.40 zu erwarten. Eine MIC < .10 erhöht dagegen die Wahrscheinlichkeit, dass die Itembeziehungen eine Komplexität aufweisen, die nicht durch eine einzelne Dimension erklärt werden kann. Eine MIC > .50 lässt indes vermuten, dass die in der Skala zusammengefassten Items eine erhöhte Redundanz aufweisen und die Messung des interessierenden Konstrukts zu spezifisch ausfällt. Nach STREINER kann auch ein *Cronbach- α* > .90 als ein Hinweis auf Redundanz und weniger auf Homogenität gewertet werden. (Briggs/Cheek (1986), S. 115; Bühner (2006), S. 133; Bühner (2011), S. 168; Moosbrugger/Kelava (2012), S. 133f; Streiner (2003), S. 102)

Die Präzision von α kann als Standardfehler von *Cronbach- α* verstanden werden und gibt an, wie stark die Korrelationen zwischen den Items der Skala voneinander abweichen. Je höher der Wert von P_α , desto heterogener sind die Korrelationen, und desto eher kann Mehrdimensionalität für die fragliche Skala angenommen werden. Nach CORTINA muss die Präzision von α sehr kleine Werte annehmen ($P_\alpha < .01$), um als Anzeichen für Eindimensionalität angesehen werden zu können. (Bühner (2006), S. 133; Bühner (2011), S. 168; Cortina (1993), S. 100ff; Moosbrugger/Kelava (2012), S. 133f)

Des Weiteren muss beachtet werden, in welchem Bereich die Item-Korrelationen streuen und wie groß die Varianz der Item-Korrelationen ist. Liegt ein Teil des Streubereichs im negativen Bereich, so kann man davon ausgehen, dass die Skala keine Eindimensionalität aufweist. Eine niedrige Streuung der Korrelationswerte kann dagegen ein Hinweis auf Eindimensionalität sein. (Briggs/Cheek (1986), S. 114; Bühner (2011), S. 243; Cortina (1993), S. 100)

Ein Hinweis auf Eindimensionalität der Skala ergibt sich demnach dann, wenn die mittlere Inter-Item-Korrelation angemessen hoch und die Präzision von α sehr niedrig ist. Gleichzeitig sollte der Streubereich der Item-Korrelationen vollständig im positiven Bereich liegen und die Varianz der Korrelationen niedrig sein. Ob tatsächlich Eindimensionalität gegeben ist, kann in letzter Instanz allerdings erst durch eine Faktorenanalyse bestimmt werden. (Bühner (2006), S. 133; Bühner (2011), S. 168; Moosbrugger/Kelava (2012), S. 133f)

Tab. 6.14: Inter-Item-Korrelationen und Reliabilitätsstatistiken (N = 220)

	MIC ^a	Min ^b	Max ^b	σ^2_{MIC} ^c	σ_{MIC} ^d	α^e	α_{st}^f	P_α^g	n_{Item}^h
Iteration 1	,195	-,039	,416	,008	,0894	,803	,805	,0077	17
Iteration 2	,206	,013	,416	,007	,0837	,805	,806	,0077	16
Iteration 3	,215	,013	,416	,007	,0837	,804	,804	,0082	15
Iteration 4	,225	,047	,416	,007	,0837	,801	,802	,0088	14

^a Mean-Inter-Item-Korrelation bzw. mittlere Inter-Item-Korrelation (Brigg/Cheek (1986), S. 114f)

^b Wertebereich innerhalb dem die Inter-Item-Korrelationen streuen (Brigg/Cheek (1986), S. 114)

^c Varianz der Inter-Item-Korrelationen

^d Standardabweichung der Inter-Item-Korrelationen

^e Cronbach- α

^f Cronbach- α für standardisierte Items

^g Präzision von Cronbach- α (Cortina (1993), S. 100; Bühner (2006), S. 133, S. 145)

^h Anzahl der Items

Quelle: eigene Berechnung

Iteration 1 beinhaltet sämtliche Items, die infolge der Itemselektion nach der Schwierigkeits- und Trennschärfeanalyse zur Objektiv-Skala zusammengefasst wurden. Durch den Ausschluss der 25 Items mit niedrigen Trennschärfen und/oder ungünstigen Itemschwierigkeiten hat sich die Varianz der Skala aus den verbleibenden Items und damit auch deren Trennschärfen verändert. Im Zuge der Alpha-Maximierung sollen nun schrittweise diejenigen Items aus dieser Skala entfernt werden, die niedrige Trennschärfen aufweisen. (Bühner (2011), S. 249)

Zunächst ist festzuhalten, dass bereits die Skala in *Iteration 1* eine Reliabilität aufweist, die als gut ($\alpha > ,80$) bezeichnet werden kann. Die Skala ist somit ausreichend reliabel. Allerdings weisen innerhalb von *Iteration 1* sowohl Item v2.8 als auch Item v2.40 Trennschärfen im niedrigen Bereich ($r_{i(x-i)} < ,30$) auf. Dabei würde nur der Ausschluss von v2.8 auch eine leichte Reliabilitätssteigerung ($\alpha = ,803 \rightarrow \alpha = ,805$) bewirken. Entfernt man hingegen v2.40 aus der Skala würde sich deren Reliabilität geringfügig verringern ($\alpha = ,803 \rightarrow \alpha = ,801$). Ein weiterer Hinweis darauf, dass es sich bei der 17 Items umfassenden Skala in *Iteration 1* noch nicht um die optimale Itemauswahl zur Abbildung Generalisierten Vertrauens handelt, ergibt sich aus der mittleren Inter-Item-Korrelation mit MIC = .195 sowie dem Bereich, innerhalb dem die Itemkorrelationen (Min = -.039, Max = .416) streuen. Ausgehend von diesen Werten weist die Skala noch keine optimale Homogenität auf. So korreliert zumindest ein geringer Prozentsatz der Items negativ miteinander, sodass möglicherweise eine punktuelle Abweichung von der einheitlichen Messrichtung Generalisiertes Vertrauen vorliegt. Zusätzlich ist zu erwarten, dass sich die Komplexität der Itemstruktur nicht durch eine einzelne Dimension abbilden lässt. Dagegen streuen die Itemkorrelationen nur schwach und die Präzision von α ist niedrig, was wiederum einen Hinweis auf Eindimensionalität anzeigt. (Briggs/Cheek (1986), S. 114f; Bühner (2006), S. 133, S. 144f; Bühner (2011), S. 243)

In *Iteration 2* wurde die Skala unter Ausschluss von v2.8 erneut geprüft. Die Reliabilität konnte leicht gesteigert werden und bewegt sich nach wie vor in einem für „gut“ befundenen Bereich ($\alpha > ,80$). Auch in *Iteration 2* weist Item v2.40 eine niedrige Trennschärfe ($r_{i(x-i)} < ,30$) auf, was auf eine geringe Item-Skalenpassung hindeutet. Allerdings bewirkt ein Ausschluss dieses Items keine weitere Reliabilitätssteigerung, sondern vielmehr eine marginale Verringerung der Skalenreliabilität ($\alpha = ,805 \rightarrow \alpha = ,804$). Da bei einer Eliminierung von v2.40 nur eine geringe Reduktion der Reliabilität zu erwarten ist, scheint dieses Item nur wenig zur Messung Generalisierten Vertrauens beizutragen und sollte daher entfernt werden. Der Wert der MIC hat sich verglichen mit *Iteration 1* verbessert und liegt nun im von BRIGGS/CHEEK für optimale Homogenität angenommenen Bereich (MIC = .20-.40). Betrachtet man den Streubereich der Itemkorrelationen, so liegt dieser vollkommen in einem positiven Wertebereich. Die Streuung

6. Pilotstudie

der Itemkorrelationen hat sich verringert und die Präzision von α ist niedrig ($P_{\alpha} < .01$). Zusammengenommen stellen diese Ergebnisse einen starken Hinweis auf eine eindimensionale Skala dar.

Tab. 6.15: Item-Skala-Statistiken (Iteration 1 und 2; N = 220)

Item	Iteration 1			Iteration 2		
	$r_{i(x-i)}^a$	$R^2_{\text{mult}}^b$	α_{n-i}^c	$r_{i(x-i)}^a$	$R^2_{\text{mult}}^b$	α_{n-i}^c
v2.6	,414**	,293	,792	,419**	,291	,794
v2.7	,392**	,273	,793	,401**	,273	,795
v2.8	,217**	,171	,805	---	---	---
v2.10	,330**	,187	,797	,349**	,178	,799
v2.11	,406**	,256	,793	,384**	,231	,796
v2.13	,427**	,241	,791	,418**	,236	,793
v2.14	,363**	,229	,795	,376**	,211	,797
v2.15	,461**	,317	,788	,472**	,316	,789
v2.17	,323**	,168	,798	,318**	,162	,801
v2.23	,414**	,267	,792	,409**	,256	,794
v2.24	,389**	,241	,793	,372**	,220	,797
v2.32	,508**	,339	,786	,521**	,335	,787
v2.37	,401**	,267	,793	,409**	,265	,794
v2.39	,318**	,218	,798	,312**	,218	,800
v2.40	,291**	,181	,801	,278**	,166	,804
v2.41	,473**	,289	,787	,473**	,284	,789
v2.42	,582**	,387	,779	,584**	,385	,781

^a Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^b Quadrierte multiple Korrelation

^c Cronbach- α , wenn Item gelöscht

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

Quelle: eigene Berechnung

Für *Iteration 3* wurde nun auch Item v2.40 aus der Skala entfernt. Wie erwartet hat sich die Reliabilität der Skala leicht verringert, liegt aber immer noch oberhalb von $\alpha = .80$. Interessant ist, dass sich in *Iteration 3* die Werte des unstandardisierten Cronbach- α und die Werte des standardisierten Cronbach- α aneinander angleichen ($\alpha = .804 \leftrightarrow \alpha_{st} = .804$). Alle Items der Skala weisen demnach gleiche Varianzen auf. (Bühner (2011), S. 241) Item v2.39 erreicht lediglich eine niedrige Trennschärfe ($r_{i(x-i)} < .30$), sodass für dieses Item zu erwarten ist, dass es die Zieldimension „Generalisiertes Vertrauen“ nur unzureichend widerspiegelt. Es sollte daher ausgeschlossen werden. Wie bereits in der vorherigen Iteration würde eine Elimination dieses Items eine leichte Verringerung der Skalenreliabilität bewirken ($\alpha = .804 \rightarrow \alpha = .801$). Da es sich auch hier um einen überschaubaren Rückgang der Reliabilität handelt, ist der Ausschluss des Items vertretbar. Die mittlere Inter-Item-Korrelation hat sich gesteigert und ist nach wie vor im Bereich für optimale Homogenität angesiedelt (MIC = .20-.40). Der Streubereich der Itemkorrelationen ist im Vergleich zu *Iteration 2* unverändert, Gleiches gilt für die Varianz der Itemkorrelationen. Die Präzision von α ist mit $P_{\alpha} = .0082$ dagegen leicht angestiegen, liegt aber immer noch unterhalb des Schwellenwertes. Insgesamt kann damit auch für die Skala aus *Iteration 3* weitestgehend Homogenität und Eindimensionalität angenommen werden.

Iteration 4 führt eine erneute Reliabilitätsschätzung für die Objektiv-Skala ohne die Items v2.8, v2.39 und v2.40 durch. Die Reliabilität hat sich etwas verringert, ist mit $\alpha = .801$ aber als „gut“ zu bewerten. Alle Items weisen Trennschärfen im mittleren ($r_{i(x-i)} = .30$ -.50) oder hohen ($r_{i(x-i)} > .50$) Bereich auf, alle Items besitzen somit eine angemessene Skalenpassung und bilden

6. Pilotstudie

die Zieldimension ausreichend trennscharf ab. Die mittlere Inter-Item-Korrelation hat sich verglichen mit Iteration 3 weiter gesteigert ($MIC = .215 \rightarrow MIC = .225$), liegt aber weiterhin in dem Wertebereich, für den eine optimale Homogenität erwartet werden kann. Auch der Streubereich der Itemkorrelationen hat sich verringert, während die Varianz gleichgeblieben ist. Insgesamt scheinen die Korrelationswerte aber leicht zusammengedrückt zu sein. Die Präzision von α hat sich dagegen erhöht, liegt aber noch unterhalb des Schwellenwertes von $P_\alpha = .01$. Insgesamt scheint sich somit eine vergleichsweise homogene und eher eindimensionale Skala ergeben zu haben. Sie enthält ausschließlich trennscharfe Items und weist eine „gute“ Reliabilität auf.

Tab. 6.16: Item-Skala-Statistiken (Iteration 3 und 4; N =220)

Item	Iteration 3			Iteration 4		
	$r_{i(x-j)}^a$	$R^2_{mult}^b$	α_{n-i}^c	$r_{i(x-j)}^a$	$R^2_{mult}^b$	α_{n-i}^c
v2.6	,416**	,287	,793	,421**	,286	,789
v2.7	,399**	,273	,794	,400**	,269	,791
v2.8	---	---	---	---	---	---
v2.10	,343**	,174	,798	,340**	,170	,796
v2.11	,380**	,227	,795	,363**	,201	,793
v2.13	,426**	,236	,792	,424**	,235	,789
v2.14	,367**	,200	,797	,368**	,200	,793
v2.15	,472**	,314	,788	,475**	,314	,784
v2.17	,329**	,158	,800	,330**	,158	,797
v2.23	,409**	,253	,793	,424**	,244	,789
v2.24	,380**	,218	,795	,369**	,205	,793
v2.32	,536**	,332	,784	,546**	,330	,779
v2.37	,418**	,263	,793	,402**	,247	,791
v2.39	,285**	,170	,801	---	---	---
v2.40	---	---	---	---	---	---
v2.41	,461**	,269	,789	,464**	,269	,785
v2.42	,591**	,385	,779	,583**	,374	,775

^a Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^b Quadrierte multiple Korrelation

^c Cronbach- α , wenn Item gelöscht

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

Quelle: eigene Berechnung

Die Schätzung der inneren Konsistenz reduziert die aus der Schwierigkeits- und Trennschärfeanalyse resultierende Objektiv-Skala um weitere 3 Items auf nun insgesamt 14 Items. Ausgeschlossen wurden die Items v2.8, v2.39 und v2.40 aufgrund niedriger Trennschärfe und einer deshalb zu erwartenden gering ausgeprägten Item-Skala-Passung. Im Zuge der Elimination dieser Items konnte außerdem die Homogenität der Skala verbessert werden, sodass nun anzunehmen ist, dass die Skala eindimensional ist. Die resultierende Skala setzt sich nun aus den folgenden Items zusammen:

Tab. 6.17: Resultat der Reliabilitätsanalyse für Objektiv-Skala

(1)	2.6) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen. (-)
(2)	2.7) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken. (-)
(3)	2.10) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinandertreffen, ist von andauernden Konflikten geprägt. (-)
(4)	2.11) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden. (+)
(5)	2.13) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen. (+)
(6)	2.14) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen. (-)
(7)	2.15) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt. (-)
(8)	2.17) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte. (-)
(9)	2.23) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden. (-)
(10)	2.24) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist. (+)
(11)	2.32) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun. (-)
(12)	2.37) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden. (+)
(13)	2.41) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt. (-)
(14)	2.42) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern. (-)

Hoher Itemscore entspricht starker Ausprägung der Zieldimension, niedriger Itemscore entspricht schwacher Ausprägung der Zieldimension. Es ergeben sich daher die folgenden Codierungen für die Antwortskala:

(+) Itempolung positiv: „Stimme nicht zu“ (0); „Stimme eher zu“ (1); „Stimme teils zu, stimme teils nicht zu“ (2); „Stimme eher zu“ (3); „Stimme zu“ (4)

(-) Itempolung negativ: „Stimme nicht zu“ (4); „Stimme eher zu“ (3); „Stimme teils zu, stimme teils nicht zu“ (2); „Stimme eher zu“ (1); „Stimme zu“ (0)

6.2.2.2 Split-Half-Reliabilität

Wie bereits beschrieben, wird zur Schätzung der Split-Half-Reliabilität der zu prüfenden Test in zwei möglichst parallele Hälften zerlegt und die Korrelation der beiden Testhälften bestimmt. Diese Korrelation wird als Reliabilität des halben Tests interpretiert und dient als Grundlage zur Abschätzung der Reliabilität des gesamten Tests. Dazu wird dieser sogenannte *Halbtest-Reliabilitätskoeffizient* mit Hilfe der Spearman-Brown-Formel bzw. der Guttman-Formel auf die Gesamtlänge des Tests aufgewertet. (Bühner (2011), S. 161ff; S. 236; Lienert/Raatz (1998), S. 182ff; Moosbrugger/Kelava (2012), S. 128ff)

Der *Spearman-Brown-Koeffizient* nimmt dabei an, dass es sich bei den beiden Testhälften um zwei exakt parallele Messmethoden mit gleichen Varianzen handelt. Ist diese Annahme allerdings verletzt, kann es zu einer Über- oder Unterschätzung der wahren Reliabilität kommen. Da sich in der Realität allerdings kaum exakt parallele Testhälften finden, wird zur Schätzung der Reliabilität auf der Basis von Testhälften mit unterschiedlichen Standardabweichungen auch der *Guttman-Split-Half-Koeffizient* verwendet. Dieser Koeffizient liefert eine konservativere Schätzung der Reliabilität. (Bühner (2011), S. 161ff; S. 236; Kristof (1983), S. 557; Lienert/Raatz (1998), S. 182ff; Moosbrugger/Kelava (2012), S. 128ff)

Zunächst muss die Objektiv-Skala in zwei Testhälften unterteilt werden. Um Abweichungen hinsichtlich der Übereinstimmung bzw. Parallelität der Testhälften auszugleichen, werden sowohl die Odd-Even-Methode als auch die Methode der Itemzwillinge zur Erstellung der Testhälften eingesetzt. Danach wird für jede der so ermittelten Halbtest-Kombinationen eine Reliabilitätsschätzung durchgeführt und die Ergebnisse verglichen. (Bühner (2011), S. 157; Lienert/Raatz (1998), S. 183f; Moosbrugger/Kelava (2012), S. 129f)

6. Pilotstudie

Zur Bestimmung der Zusammensetzung der Testhälften mit der Odd-Even-Methode wurden den Items der Objektiv-Skala neue Reihungsnummern von 1 bis 14 entsprechend ihrer ursprünglichen Reihung zugewiesen. Danach wurden alle Items mit ungerader Reihungsnummer *Testhälfte 1* und alle Items mit gerader Reihungsnummer *Testhälfte 2* zugeordnet. Es ergibt sich somit die folgende Zusammensetzung für die resultierenden Testhälften:

Tab. 6.18: Testhalbierung: Odd-Even

<i>Testhälfte 1</i>		<i>Testhälfte 2</i>	
Reihungsnummer	Item	Reihungsnummer	Item
1	v2.6	2	v2.7
3	v2.10	4	v2.11
5	v2.13	6	v2.14
7	v2.15	8	v2.17
9	v2.23	10	v2.24
11	v2.32	12	v2.37
13	v2.41	14	v2.42

Zur Erstellung der Testhälften mit der Methode der Itemzwillinge wurden in einem ersten Schritt Itempaarungen gebildet, bei denen Itemschwierigkeit, Itemstreuung und Trennschärfe möglichst ähnlich sind. In einigen Fällen konnten allerdings nur für jeweils zwei dieser Parameter Passungen gefunden werden. In einem zweiten Schritt wurden die Itempaarungen getrennt und jeweils per Zufall auf die Testhälften verteilt. Es resultieren daraus die folgenden Testhälften:

Tab. 6.19: Testhalbierung: Itemzwillinge

<i>Testhälfte 1</i>					<i>Testhälfte 2</i>				
Item	P _i ^a	P _m ^b	σ _i ^c	r _{i (x-i)} ^d	Item	P _i ^a	P _m ^b	σ _i ^c	r _{i (x-i)} ^d
v2.15	3,5682	,642	1,11850	,475**	v2.37	3,7500	,688	,85808	,402**
v2.24	3,3955	,599	,92306	,369**	v2.11	3,0636	,516	,80837	,363**
v2.23	3,0545	,514	,99622	,424**	v2.13	3,0682	,517	,91656	,424**
v2.10	2,9091	,477	1,01626	,340**	v2.42	2,8636	,466	1,02460	,583**
v2.41	2,5091	,377	1,04024	,464**	v2.32	2,2136	,303	,90397	,546**
v2.14	2,2773	,319	1,00702	,368**	v2.17	2,5545	,389	1,05628	,330**
v2.6	2,0818	,270	,81237	,421**	v2.7	2,2182	,305	,86939	,400**

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

*p<.05 (einseitig), **p<.01 (einseitig)

(Quelle: eigene Berechnungen)

Die Ergebnisse für die Odd-Even-Methode zeigen, dass die beiden Testhälften jeweils eher niedrige Reliabilitäten ($\alpha < .70$) aufweisen. Außerdem scheinen die Testhälften nicht exakt parallel zu verlaufen, da *Testhälfte 1* eine etwas höhere innere Konsistenz als *Testhälfte 2* aufweist. Beide Testhälften korrelieren allerdings stark miteinander. Die Schätzung der Reliabilität der Gesamtskala als Aufwertung der Halbttest-Reliabilität ergibt $r_{tt} = .799$ bzw. $\rho_{tt} = .797$. Beide Werte sind mit „akzeptabel“ zu bewerten, wobei der *Guttman-Split-Half-Koeffizient* aufgrund der nicht gegebenen Parallelität der Testhälften die aussagekräftigere Schätzung darstellt.

Die Schätzung der Split-Half-Reliabilität auf der Basis von Itemzwillingen ergibt ebenfalls Unterschiede in den Reliabilitäten der beiden Testhälften. So weist *Testhälfte 1* eine etwas niedrigere innere Konsistenz auf als *Testhälfte 2*. *Cronbach-α* nimmt aber in beiden Fällen eher niedrige Werte ($\alpha < .70$) an. Beide Testhälften korrelieren allerdings stark miteinander. Für die

6. Pilotstudie

Schätzung der Reliabilität der Gesamtskala ergeben sich $r_{tt} = .818$ bzw. $\rho_{tt} = .818$. Diese Werte sind jeweils als „gut“ zu bewerten.

Tab. 6.20: Reliabilitätsstatistiken

Testhalbierungsmethode	α_1^a	α_2^b	r_{12}^c	r_{tt}^d	ρ_{tt}^e
Odd-Even	,694	,641	,666	,799	,797
Itemzwillinge	,643	,680	,693	,818	,818

^a Cronbach- α für Testhälfte 1

^b Cronbach- α für Testhälfte 2

^c Korrelation zwischen den Testhälften

^d Spearman-Brown-Koeffizient (Lienert/Raatz (1998), S. 185; Bühner (2011), S. 162; Moosbrugger/Kelava (2012), S. 128f)

^e Guttman-Split-Half-Koeffizient (Lienert/Raatz (1998), S. 186; Bühner (2011), S. 164)

Vergleicht man die auf Basis der verschiedenen Testhalbierungsverfahren ermittelten Split-Half-Reliabilitäten ergeben sich nur marginale Unterschiede. Die auf Basis der Odd-Even-Methode ermittelte Reliabilität der Gesamtskala ist etwas niedriger als die mit Hilfe von Itemzwillingen ermittelte Reliabilität. Sie liegen damit entweder leicht unter oder leicht über der Schätzung der Reliabilität der Gesamtskala mittels Cronbach- α ($\alpha = .801$). Insgesamt scheint die Objektiv-Skala somit eine in der Tendenz mit „gut“ zu bewertende Reliabilität aufzuweisen.

6.2.3 Prüfung auf Eindimensionalität (Objektiv-Skala)

Im Rahmen der Reliabilitätsanalyse wurden bisher lediglich Hinweise auf eine Eindimensionalität der Objektiv-Skala gefunden. Im Folgenden soll nun mit Hilfe strukturentdeckender Verfahren verifiziert werden, ob es sich um eine mehrdimensionale Skala handelt oder ob die Skala tatsächlich nur eine einzelne Dimension abbildet. (Mummendey/Grau (2014), S. 129ff)

Es kann prinzipiell davon ausgegangen werden, dass eine Messung immer messfehlerbehaftet ist und deshalb die Varianz innerhalb der Items nie vollständig durch ein zugrunde gelegtes Modell aufgeklärt werden kann. Vor diesem Hintergrund wurde entschieden, zur Prüfung der Eindimensionalität der Objektiv-Skala eine Hauptachsenanalyse durchzuführen. Ziel dieses Verfahrens ist es, „latente Konstrukte bzw. Faktoren zu identifizieren, welche das Beziehungsmuster zwischen den Variablen erklären, wobei nicht die gesamte Varianz der beobachteten Variablen, sondern nur die wahre Varianz durch zugrunde liegende gemeinsame Faktoren erklärt werden kann“ (Moosbrugger/Kelava (2012), S. 328). Leitend bei der Interpretation der extrahierten Faktoren ist dabei die folgende Frage:

„Wie lässt sich die Ursache bezeichnen, die für die hohen Korrelationen zwischen den Variablen verantwortlich sind?“ (Backhaus/Erichson/Plinke/Weiber (2006), S. 351)

Es handelt sich hierbei um eine explorative Faktorenanalyse, die auf der Annahme beruht, dass sich die Varianz der in die Analyse eingehenden Variablen in einen durch latente Faktoren verursachten systematischen Varianzanteil und einen Fehlervarianzanteil zerlegen lässt. Der durch die Faktoren aufgeklärte Varianzanteil wird im Rahmen dieses Verfahrens als *Kommunalität* bezeichnet, während der Fehlervarianzanteil als Residualvarianz oder *Uniqueness*⁴⁸ bezeichnet wird. Die *Uniqueness* wird nochmals in die Spezifität und den Messfehler eines Items differenziert. Dabei beschreibt die Spezifität den systematischen Varianzanteil, der nicht durch andere Faktoren erklärt werden kann, während der Messfehler

⁴⁸ Vgl. Bühner (2011), S. 300

den Anteil der Fehlervarianz enthält, der weder auf die Faktoren noch auf andere Variablen zurückgeführt werden kann. Die mit Hilfe der Hauptachsenanalyse extrahierten Faktoren erklären demnach nur die Varianz, die ein Item mit allen anderen Items teilt. (Backhaus/Erichson/Plinke/Weiber (2006), S. 350f; Bühner (2011), S. 300, S. 310ff; Klopp (2013), S. 1f; Moosbrugger/Kelava (2012), S. 328)

6.2.3.1 Stichprobeneignung bzw. Opportunität der faktorenanalytischen Untersuchung

Als Gütekriterien für die Opportunität einer Hauptachsenanalyse wurden die Fallzahl, das Kaiser-Meyer-Olkin-Kriterium (KMO), der Bartlett-Test, die MSA-Koeffizienten (*Measure of Sample Adequacy*) und die Anti-Image-Kovarianzmatrix herangezogen.

Hinsichtlich der für die Durchführung einer explorativen Faktorenanalyse angemessenen Fallzahl herrscht innerhalb der Literatur kein allgemeingültiger Konsens. So empfehlen GORSUCH und KLINE jeweils ein Minimum von 100 Fällen, während nach GUILFORD mindestens 200 Fälle notwendig sind und CATTELL sich für wenigstens 250 Fälle ausspricht. COMREY/LEE schlagen eine verhältnismäßig differenzierte Bewertungsskala für die Eignung der Stichprobengröße vor: 100 = *poor*, 200 = *fair*, 300 = *good*, 500 = *very good*, 1000 und mehr = *excellent*. Wenn möglich sollten allerdings immer Stichproben mit mindestens 500 Fällen genutzt werden. BRYANT/YARNOLD leiten die notwendige Fallzahl aus der Relation von Variablen zu Fällen ab. Als hinreichend erwiesen sich dabei Verhältnisse von 1:3 bis 1:10. Nach MACCALLUM/WIDAMAN/ZHANG/HONG hängt die für eine Faktorenanalyse ausreichende Fallanzahl eng zusammen mit der Itemreliabilität und der Überbestimmtheit der Faktoren. Als Mindestschätzung der Reliabilität wird dabei die Kommunalität eines Items verwendet. Nach Möglichkeit sollte dies bei sehr kleinen Stichproben ($N = 60$) mindestens $h^2 = .60$ sein. Nehmen die Kommunalitäten der Items Werte von $h^2 \approx .50$ an, so sollte die Stichprobe mindestens 100 bis 200 Fälle enthalten. Erreichen die Kommunalitäten vornehmlich Werte von $h^2 < .50$ sollte die Anzahl der extrahierten Faktoren möglichst gering sein und jeder der Faktoren durch eine möglichst hohe Zahl von Indikatoren bestimmt sein (6 oder mehr Items pro Faktor). (Bryant/Yarnold (1994), S. 100; Bühner (2011), S. 344f; Cattell (1978), S. 509; Gorsuch (1983), S. 332; Guilford (1954), S. 533; Kline (1979), S. 38ff; MacCallum/Widaman/Zhang/Hong (1999), S. 96; Mundfrom/Shaw/Ke (2005), S. 163ff; Velicer/Fava (1998), S. 231f)

Die für die hier durchgeführte Hauptachsenanalyse zugrunde gelegte Fallzahl von $N=220$ kann nach COMREY/LEE als „*fair*“ bewertet werden und ist damit geeignet für die Durchführung einer explorativen Faktorenanalyse. Auch die Kriterien von BRYANT/ARNOLD sind mit einer Relation von 14 Items zu 220 Fällen (Verhältnis von 1:15) mehr als erfüllt. Die Kriterien von MACCALLUM/WIDAMAN/ZHANG/HONG sind dagegen nur bedingt erfüllt. So ergeben sich für die anvisierte einfaktorielle Lösung lediglich Kommunalitäten von $h^2 = .10-.50$. Dabei nimmt die Mehrzahl der Kommunalitäten (ca. 64%) allerdings Wert von $h^2 > .20$ an. Nach MUNDFROM/SHAW/KE gilt bei niedrigen Kommunalitäten ($h^2 = .20-.40$) und 8 Items pro Faktor bereits eine Fallzahl von $N = 130$ als „ausgezeichnet“. In der hier durchgeführten Analyse kommen allerdings sogar 14 Items auf einen Faktor bei 220 Fällen. Insgesamt scheint die Fallzahl von $N = 220$ ausreichend groß für die Durchführung einer Hauptachsenanalyse zu sein. (Bühner (2011), S. 345; Eid/Gollwitzer/Schmitt (2010), S. 918ff; Klopp (2013), S. 4; MacCallum/Widaman/Zhang/Hong (1999), S. 84; Mundfrom/Shaw/Ke (2005), S. 163ff)

6. Pilotstudie

Das KMO gibt Hinweise darauf, ob die Itemauswahl für eine Faktorenanalyse geeignet ist. Es wird dabei die gemeinsame Varianz der Items in Relation zur Summe aus der gemeinsamen Varianz und der Summe der quadrierten Partialkorrelationen gesetzt. Je höher die quadrierten Partialkorrelationen, desto größer ist der Anteil an spezifischer Varianz, den die Items nicht mit den anderen Items teilen, und desto niedriger wird der Wert des Kaiser-Meyer-Olkin-Kriteriums. Ein hohes KMO zeigt somit an, ob substantielle Korrelationen vorliegen, die eine Faktorenanalyse rechtfertigen. Das KMO erreicht mit .844 einen Wert, der eine „gute“ Eignung der Items anzeigt. Auch die MSA-Koeffizienten liegen mit durchweg $MSA \geq .80$ im für Faktorenanalysen wünschenswerten Bereich. Das *Measure of Sample Adequacy* macht dabei eine Aussage über die Eignung der einzelnen Items bzw. darüber, wie gut ein Item zu den übrigen Items passt. Hierzu werden die Korrelationen der Items mit den restlichen Items betrachtet. Es kann somit davon ausgegangen werden, dass die Items untereinander stark korrelieren. (Backhaus/Erichson/Plinke/Weiber (2006), S. 336f; Bühner (2011), S. 346f, S. 355; Kaiser/Rice (1974), S. 111ff; Klopp (2013), S. 3f)

Der Bartlett-Test auf Sphärizität prüft, ob die Stichprobe aus einer Grundgesamtheit stammt, in der die Items unkorreliert sind. Für die hier analysierte Stichprobe ergibt sich ein hochsignifikantes Ergebnis, was darauf hindeutet, dass die Ausgangsvariablen miteinander korrelieren bzw. dass die Korrelationen der in die Analyse eingehenden Items signifikant von null abweichen. Die Items sind demnach für die Durchführung einer Faktorenanalyse geeignet. (Backhaus/Erichson/Plinke/Weiber (2006), S. 335; Bühner (2011), S. 347f, S. 355)

Tab. 6.21: Maße zur Bestimmung der Stichprobeneignung für EFA^a

N ^b	KMO ^c	MSA ^d	Bartlett-Test ^e		Anti-Image ^f
220	,844	> .80	Ungefähres χ^2	561,421	22% ungleich Null
			df	91	
			Signifikanz	,000**	

^a Explorative Faktorenanalyse

^b Fallanzahl

^c Maß der Stichprobeneignung nach KAISER-MEYER-OLKIN (Backhaus/Erichson/Plinke/Weiber (2006), S. 336; Bühner (2011), S. 346f)

^d Wertebereich der MSA-Koeffizienten (*Measure of Sample Adequacy*) (Backhaus/Erichson/Plinke/Weiber (2006), S. 336; Bühner (2011), S. 347, S. 356; Klopp (2013), S. 3f)

^e Bartlett-Test auf Sphärizität (Backhaus/Erichson/Plinke/Weiber (2006), S. 335; Bühner (2011), S. 347f)

^f Anteil der Nicht-diagonalen-Elemente ungleich Null (d.h. > .09) in der Anti-Image-Kovarianzmatrix (Backhaus/Erichson/Plinke/Weiber (2006), S. 335f; Bühner (2011), S. 356f)

*p<.05 (einseitig), **p<.01 (einseitig)

(Quelle: eigene Berechnungen)

Als Image wird der Varianzanteil eines Items bezeichnet, der durch eine multiple Regression der restlichen Items aufgeklärt werden kann. Das Anti-Image beschreibt dann den verbleibenden Varianzanteil, der von den übrigen Items unabhängig ist. Die Faktorenanalyse nimmt an, dass der Zusammenhang der Items durch gemeinsame Faktoren erklärt wird, weshalb das Anti-Image der Varianzen möglichst gering ausfallen muss. Nach DZIUBAN/SHIRKEY ist dies gegeben, wenn der Anteil der Nicht-diagonalen-Elemente der Anti-Image-Kovarianz-Matrix, die ungleich null sind (> .09) einen Schwellenwert von 25% unterschreitet. Auch dieses Kriterium wird von den hier untersuchten Items erfüllt. So erreichen nur rund 22% der Nicht-diagonalen-Elemente Werte ungleich null. (Backhaus/Erichson/Plinke/Weiber (2006), S. 335f; Bühner (2011), S. 356f; Dziuban/Shirkey (1974), S. 359)

Eine Zusammenschau der verschiedenen Gütekriterien lässt den Schluss zu, dass eine Faktorenanalyse durchaus als opportun zu bewerten ist.

6.2.3.2 Extraktions- bzw. Abbruchkriterien

Um die Anzahl der zu extrahierenden Faktoren zu bestimmen, wurden verschiedene Kriterien herangezogen. In absteigender Reihenfolge ihrer Relevanzsetzung waren dies die sinnvolle Interpretation der Faktoren auf Basis der theoretischen Vorüberlegungen, der *Minimum-Average-Partial-Test* von VELICER (MAP-Test), die *Parallelanalyse* nach HORN, der *Scree-Test* nach CATTELL und das Kaiser-Guttman- bzw. Eigenwertkriterium. Mit Ausnahme des Kaiser-Guttman-Kriteriums sind die Ergebnisse der Prüfungsverfahren deckungsgleich.

Tab. 6.22: Extraktionskriterien für Faktoren

Abbruchkriterium	Anzahl der postulierten Komponenten
Theorie ^a	1 bis 6
MAP-Test ^b	1
Parallelanalyse ^c	1
Scree-Test ^d	1 bis 2
Kaiser-Guttman-Kriterium ^e	4

(Quelle: eigene Berechnungen)

Nach dem Kaiser-Guttman-Kriterium gelten solche Faktoren als relevant, die Eigenwerte größer 1 erreichen. Der Eigenwert eines Faktors ergibt sich aus der Summe der quadrierten Faktorladungen des Faktors und gibt an, wie viel Varianz innerhalb der Items durch den Faktor erklärt werden. Faktoren mit Eigenwerten größer 1 klären dabei mehr Varianz auf als ein standardisiertes Item. Mit anderen Worten: Der Faktor erklärt mehr Unterschiede im Antwortverhalten als ein einzelnes Item. In der vorliegenden Analyse ergaben sich anhand des Kaiser-Guttman-Kriteriums 4 Komponenten. Dabei ist allerdings zu bedenken, dass das Eigenwertkriterium die wirkliche Anzahl der Faktoren in der Regel überschätzt. So weisen einige der extrahierten Faktoren oft per Zufall Eigenwerte größer 1 auf. (Backhaus/Erichson/Plinke/Weiber (2006), S. 353; Bühner (2011), S. 321; Moosbrugger/Kelava (2012), S. 329f; Zwick/Velicer (1986), S. 434, S. 439)

Auf den ersten Blick ist der Scree-Plot⁴⁹ im vorliegenden Fall verhältnismäßig eindeutig zu interpretieren. Der erste „Knick“ ergibt sich nach dem ersten Faktor und die Eigenwerte nähern sich danach asymptotisch der x-Achse, was für eine einfaktorielle Lösung spricht. Außerdem ist die Differenz der Eigenwerte zwischen dem ersten und dem zweiten Faktor deutlich größer als die Differenzen zwischen den verbleibenden Faktoren. Daneben können aus dem Scree-Plot allerdings auch Hinweise auf eine mehrfaktorielle Lösung abgelesen werden. So sinken die Eigenwerte nach dem zweiten Faktor nochmals deutlich ab und der Kurvenverlauf wird flacher. Dies deutet daraufhin, dass die Faktoren 1 und 2 den Löwenanteil zur Varianzaufklärung beitragen. Eine zweifaktorielle Lösung würde zudem der Konvention entsprechen, auch den

⁴⁹ Der *Scree-Test* stellt ein gleichermaßen einfaches wie subjektives Kriterium zur Bestimmung der Faktorenanzahl dar. Hierbei werden die Eigenwerte der Faktoren nach Größe geordnet in ein Koordinatensystem eingetragen und durch eine Gerade verbunden. An der Stelle, wo die Eigenwerte einen Sprung machen bzw. wo die Differenz zwischen den Eigenwerten zweier Faktoren am größten ist, ergibt sich dann ein „Knick“ im sogenannten Scree-Plot. In der Regel nähert sich der Graph nach diesem „Knick“ asymptotisch der x-Achse, sodass durch den Punkt im „Knick“ die Anzahl der zu extrahierenden Faktoren bestimmt wird. Die Faktoren mit den niedrigsten Eigenwerten werden demnach als inhaltlich irrelevant angesehen. Hierbei können allerdings verschiedene Interpretationsschwierigkeiten entstehen: So herrscht innerhalb der Literatur keine Einigkeit darüber, ob nur diejenigen Faktoren als inhaltlich relevant gelten, die vor dem „Knick“ liegen oder ob auch der Faktor nach dem „Knick“ miteinbezogen wird. Gleichzeitig ist es aber auch möglich, dass sich kein „Knick“ ergibt, sich mehrere „Knicke“ ergeben oder keine eindeutige Linie entlang der niedrigen Eigenwerte gezogen werden kann. (Backhaus/Erichson/Plinke/Weiber (2006), S. 353f; Bühner (2011), S. 322f; Cattell (1966); Moosbrugger/Kelava (2012), S. 330; Zwick/Velicer (1986), S. 434)

Faktor nach dem „Knick“ zu berücksichtigen. Vor dem Hintergrund, dass es sich bei der Interpretation des Scree-Plot meist um eine vergleichsweise subjektive Angelegenheit handelt, wird das Ergebnis des Scree-Tests in dieser Analyse allenfalls als unterstützendes grafisches Entscheidungskriterium für objektivere Abbruchkriterien genutzt werden. (Backhaus/Erichson/Plinke/Weiber (2006), S. 353f; Bühner (2011), S. 322f; Cattell (1966); Moosbrugger/Kelava (2012), S. 330; Zwick/Velicer (1986), S. 434)

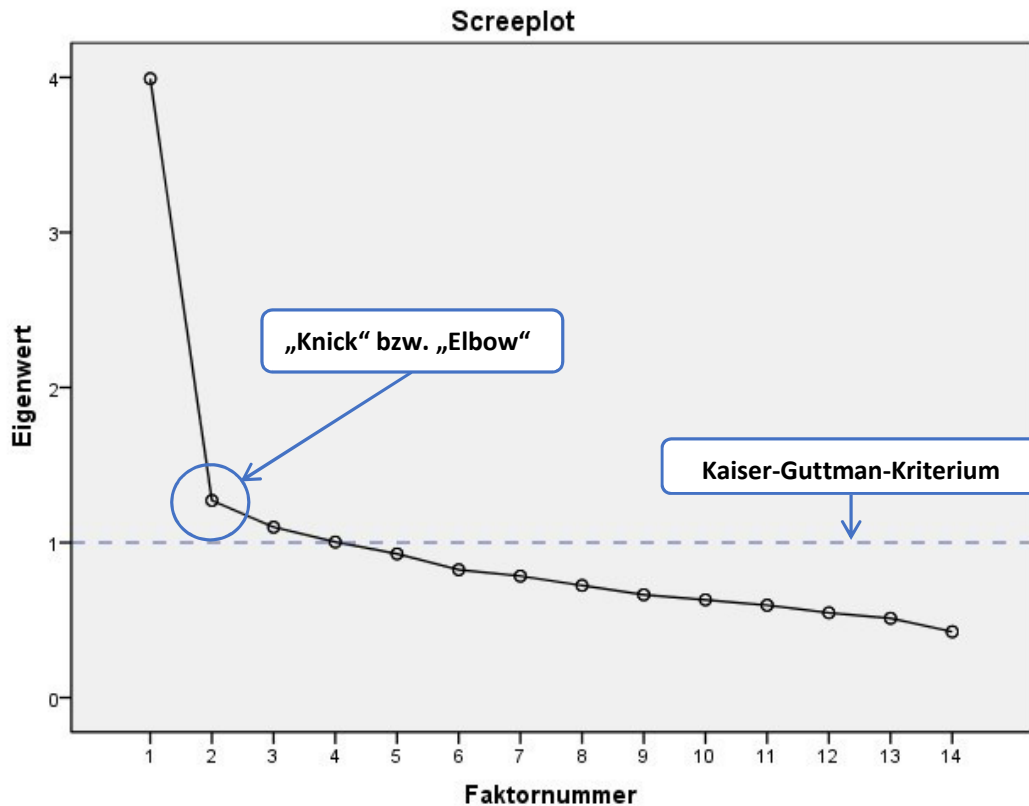


Abb. 6.15: Scree-Test und Kaiser-Guttman-Kriterium (Eigenwert > 1) für Faktorenlösung erster Ordnung (Objektiv-Skala)

Objektivere Methoden zur Schätzung der Faktorenanzahl stellen nach Bühner (2011) die Parallelanalyse nach HORN und der MAP-Test von VELICER dar.

„Als besonders geeignete Methoden, um die Faktorenanzahl zu bestimmen, haben sich der MAP-Test und die Parallelanalyse erwiesen. Wir empfehlen beide Methoden anzuwenden und dann bei einer Entscheidung, falls unterschiedliche Faktorenanzahlen resultieren, die besser interpretierbare Lösung zu verwenden.“ (Bühner (2011), S. 328)

Für die hier berechnete Parallelanalyse⁵⁰ wurden insgesamt 1000 Datensets mit je 220 Fällen und 14 Variablen simuliert. In der unten aufgetragenen Grafik finden sich der Verlauf der

⁵⁰ Bei der Parallelanalyse nach HORN werden die Eigenwerte aus den empirisch gewonnenen Daten mit denen aus normalverteilten Zufallsdaten verglichen. Konzeptuelle Grundlage der Parallelanalyse ist die Annahme, dass die empirisch gewonnenen Daten innerhalb der Population in Wahrheit unkorreliert sind und sich aus diesem Grund Scheinfaktoren ergeben. Dem empirisch gewonnenen Eigenwertverlauf wird deshalb ein Eigenwertverlauf gegenübergestellt, der sich aus zufallsgenerierten Datensets mit in der Population unkorrelierten Variablen ableitet. Mit anderen Worten: Es wird eine Vielzahl von Zufallsdatensätzen erzeugt, bei denen die Variablen rein zufällig in unterschiedlich starkem Ausmaß miteinander korrelieren. Für jedes dieser Datensets wird eine Faktorenanalyse

6. Pilotstudie

empirischen Eigenwerte (*rawdata*), der Verlauf der gemittelten zufallsgenerierten Eigenwerte (*means*) und der Verlauf der Eigenwerte, unter dem 95% der Eigenwerte aus Zufallszahlen liegen (*percnytl*). Ist der empirische Eigenwert größer als dieses 95%-Perzentil, so gehört er zu den extremsten 5% der zufallsgenerierten Eigenwerte und es ist sehr unwahrscheinlich, dass er zufällig zustande gekommen ist. Es handelt sich somit um einen systematischen Faktor. (Bühner (2011), S. 323ff; Klopp (2013), S. 5f; Moosbrugger/Kelava (2012), S. 331f; O'Connor (2000), S. 397ff; Zwick/Velicer (1986), S. 434f)

Tab. 6.23: Parallelanalyse nach HORN (Ncases = 220; Nvars = 14; Ndatsets = 1000; Percent = 95)

Root ^a	Rawdata ^b	Means ^c	Percnytl ^d
1	3,992016	1,447224	1,547774
2	1,271240	1,339787	1,410913
3	1,099504	1,259503	1,319097
4	1,003636	1,188598	1,240258
5	,927241	1,124434	1,174624
6	,824820	1,064767	1,109209
7	,784111	1,010384	1,052047
8	,723279	,955929	,995692
9	,663640	,903480	,945409
10	,630571	,851319	,893332
11	,596071	,800083	,841669
12	,546953	,746102	,793263
13	,511661	,688895	,739484
14	,425258	,619494	,676136

^a Faktor (O'Connor (2000); Bühner (2011), S. 324)

^b empirische Eigenwerte (O'Connor (2000); Bühner (2011), S. 324)

^c gemittelte zufallsgenerierte Eigenwerte (O'Connor (2000); Bühner (2011), S. 324)

^d Eigenwerte, unter denen 95% der zufallsgenerierten Eigenwerte liegen (O'Connor (2000); Bühner (2011), S. 324)

berechnet und die sich ergebenden Eigenwerte pro Faktor gemittelt. Im Rahmen der Faktorenanalyse der empirischen Daten sollen dann nur diejenigen Faktoren extrahiert werden, deren empirische Eigenwerte größer als die gemittelten Eigenwerte der zufallsgenerierten Daten sind. (Bühner (2011), S. 323ff; Klopp (2013), S. 5f; Moosbrugger/Kelava (2012), S. 331f; O'Connor (2000), S. 397ff; Zwick/Velicer (1986), S. 434f)

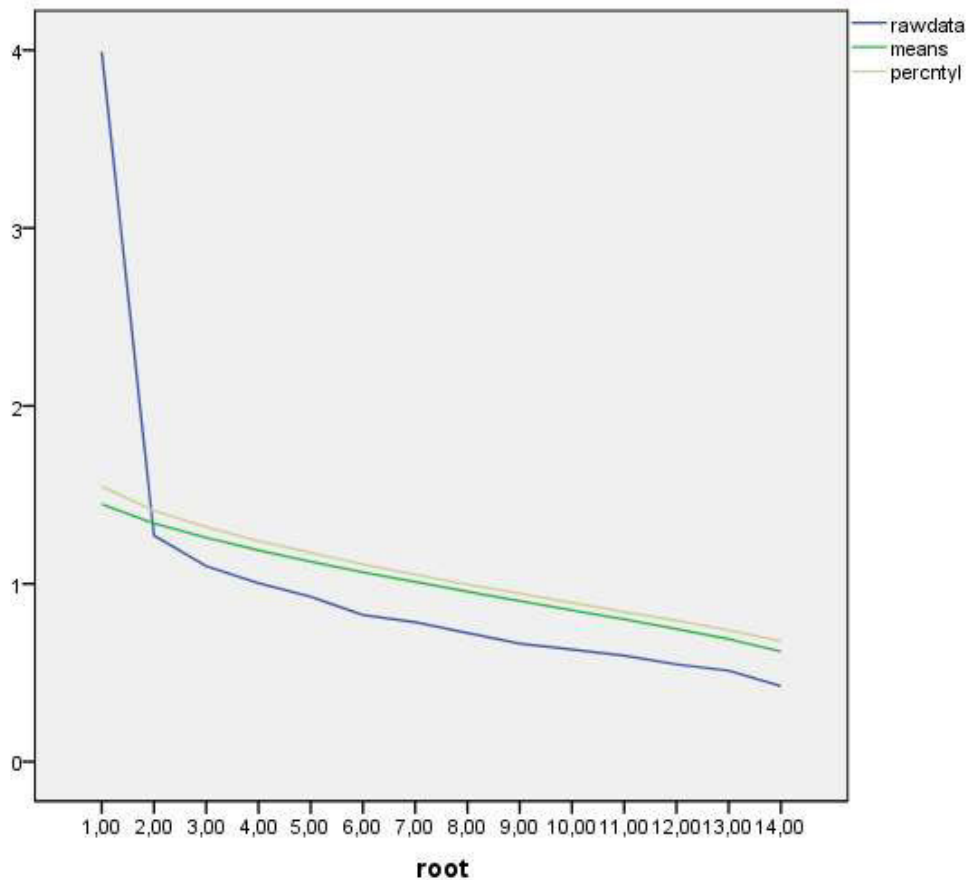


Abb. 6.16: Ergebnis der *Parallelanalyse nach Horn* mit empirisch beobachtetem Eigenwertverlauf (*rawdata*), mittlerem zufallsgeneriertem Eigenwertverlauf (*means*) und 95%-Perzentil der Verteilung zufällig erzeugter Eigenwerte (*percnyl*) für Faktoriellösung erster Ordnung (Objektiv-Skala)

Aus den Daten geht hervor, dass lediglich der Eigenwert des ersten Faktors größer als der entsprechende Eigenwert des 95%-Perzentils ist. Die Parallelanalyse empfiehlt somit eine einfaktorielle Lösung.

Im Allgemeinen überschätzt eine Parallelanalyse die Anzahl zu extrahierender Faktoren nur selten. Nach BEAUDUCEL kann es bei starken ersten Hauptkomponenten mit einem hohen Eigenwert aber zu einer Unterschätzung der wahren Faktorenzahl kommen. Allerdings gelten diese Resultate vornehmlich für Hauptkomponentenanalysen. Parallelanalysen, die im Rahmen von Hauptachsenanalysen durchgeführt werden, neigen nach BÜHNER eher zu einer Überschätzung der wahren Faktorenzahl. Es kann daher mit hoher Wahrscheinlichkeit davon ausgegangen werden, dass die hier angezeigte einfaktorielle Lösung der wahren Anzahl von Faktoren entspricht. (Beauducel (2001); Bühner (2011), S. 323ff, S. 360ff; Klopp (2013), S. 5f; Moosbrugger/Kelava (2012), S. 331f; O'Connor (2000), S. 397ff; Zwick/Velicer (1986), S. 434f, S. 440)

Aus der nachfolgenden Ergebnistabelle des MAP-Test⁵¹ geht hervor, dass die mittlere quadrierte Partialkorrelation mit der Auspartialisierung des ersten Faktors minimal wird und

⁵¹ Der *Minimum-Average-Partial-Test* von VELICER bestimmt die zu extrahierende Faktorenanzahl im Wesentlichen durch die schrittweise Auspartialisierung der über eine Hauptkomponentenanalyse erhaltenen Faktoren aus der Korrelationsmatrix der in die Analyse eingehenden Items, bis die systematische Varianz der Korrelationsmatrix minimal wird. Im Einzelnen wird hierzu zunächst die erste Hauptkomponente auspartialisiert, sodass eine sogenannte

6. Pilotstudie

danach wieder ansteigt. Die systematische Varianz der Korrelationsmatrix wird somit bereits durch einen einzigen Faktor ausgeschöpft. Der MAP-Test empfiehlt dementsprechend eine einfaktorielle Lösung.

Das Ergebnis des MAP-Tests ist besonders verlässlich, wenn die zu extrahierenden Faktoren mit acht oder mehr Items besetzt sind. Im vorliegenden Fall werden sogar 14 Items zugeteilt, sodass die Schätzung von einer einzelnen Komponente korrekt ist. (Zwick/Velicer (1986), S. 440)

Tab. 6.24: MAP-Test von VELICER (*Minimum-Average-Partial-Test*)

Auspartialisierter Faktor	$r^2_{\text{part}}^a$
0	,0570
1	,0127
2	,0183
3	,0265
4	,0363
5	,0487
6	,0672
7	,0875
8	,1222
9	,1633
10	,2256
11	,3112
12	,4801
13	1,0000

^a Quadrierte mittlere Partialkorrelation (Bühner (2011), S. 325f)

Das Wichtigste der hier aufgezählten Abbruchkriterien stellen allerdings die theoretischen Vorüberlegungen dar. So ergaben sich aus der Reliabilitätsanalyse bereits Hinweise auf die Eindimensionalität der Skala und auch auf Basis der inhaltlichen Vorarbeit wurde erwartet, dass die Skala ausschließlich das Konstrukt „Generalisiertes Vertrauen“ abbilden würde. Es existiert somit ein theoretisches Modell, welches die inhaltliche Interpretierbarkeit einer einfaktoriellen Lösung abdeckt. (Bühner (2011), S. 321, S. 349; Klopp (2013), S. 4f) Eine einfaktorielle Lösung wird auch in dem Sinne unterstützt, als dass in der Theorie sämtliche Items eine einzelne Zieldimension, nämlich „Generalisiertes Vertrauen“, abbilden sollen. Betrachtet man allerdings die im Rahmen dieser Studie gewählte Art und Weise der Operationalisierung Generalisierten Vertrauens als mehrdimensionales Konstrukt, so wird auch die Möglichkeit einer multifaktoriellen Lösung durch theoretische Vorüberlegungen getragen. Aufgrund der Tatsache, dass in die Faktorenanalyse lediglich die aus der Reliabilitätsanalyse resultierenden Items eingehen und diese im Zuge der Operationalisierung sechs verschiedenen Komponenten zugeordnet wurden, ist somit eine Faktorenlösung mit 1 bis 6 verschiedenen Dimensionen zu erwarten.

Nachfolgend sollen sowohl eine einfaktorielle als auch eine mehrfaktorielle Lösung berechnet und diskutiert werden. Die Betrachtung einer einfaktoriellen Lösung wird dabei durch die Ergebnisse der *Parallelanalyse nach HORN* und des MAP-Tests nach VELICER sowie den Erkenntnissen aus der Reliabilitätsanalyse unterstützt. Auch der Scree-Test und die theoretischen Annahmen lassen sich in Richtung einer einfaktoriellen Lösung interpretieren bzw. decken mit ihren Extraktionsempfehlungen eine entsprechende Lösungsvariante ab. Für

Residualmatrix, bestehend aus den Partialkorrelationen der Items ohne den Einfluss des ersten Faktors, resultiert. Aus den Partialkorrelationen unterhalb oder oberhalb der Diagonalen der Residualmatrix wird dann die mittlere quadrierte Partialkorrelation berechnet. Diese Schritte werden mit den weiteren Faktoren wiederholt, bis die mittlere quadrierte Partialkorrelation nicht mehr kleiner wird. Es werden dann so viele Komponenten extrahiert, wie bis zu diesem Punkt Faktoren auspartialisiert wurden. (Bühner (2011), S. 325f, S. 363f; O'Connor (2000), S. 397; Zwick/Velicer (1986), S. 434)

die Annahme der Empfehlungen der genannten Abbruchkriterien spricht außerdem, dass sich eine vergleichsweise starke erste Hauptkomponente ergibt, deren Eigenwert und damit ihr Anteil erklärter Varianz deutlich von dem der restlichen Komponenten abgesetzt ist. (Bühner (2011), S. 349)

Eine mehrfaktorielle Lösung wird dagegen vornehmlich durch das Kaiser-Guttman-Kriterium befürwortet. Allerdings lassen ebenso der Scree-Test wie auch die theoretischen Vorüberlegungen im Rahmen der Operationalisierung die Option einer mehrere Komponenten umfassenden Faktorenlösung zu. Für eine solche Faktorstruktur spricht außerdem, dass die Varianzaufklärung deutlich erhöht wird und eine differenziertere Aufgliederung des Merkmalbereichs eher der Anlage der ursprünglichen Operationalisierung der Zieldimension „Generalisiertes Vertrauen“ entsprechen würde. (Bühner (2011), S. 321)

6.2.3.3 Einfaktorielle Lösung

Als Nächstes sollen die Itemparameter der empfohlenen einfaktoriellen bzw. eindimensionalen Lösung betrachtet werden. Da nur ein einzelner Faktor extrahiert wird, kommt bei der Berechnung der Hauptachsenanalyse keine Rotationsmethode zum Einsatz. (Backhaus/Erichson/Plinke/Weiber (2006), S. 357; Bühner (2011), S. 338)

Der extrahierte Faktor erklärt durchschnittlich ca. 23,25% der Varianz in den Items. Zum Vergleich wurde eine weitere Hauptachsenanalyse berechnet, bei der den Empfehlungen des Kaiser-Guttman-Kriteriums gefolgt wurde. Es wurde dabei davon ausgegangen, dass die zu extrahierenden 4 Faktoren nicht voneinander unabhängig sind und dementsprechend die oblique Rotationstechnik „Promax“ (Kappa-Wert = 4) angewandt. Ein Vergleich mit der vierfaktoriellen Lösung zeigte, dass die Extraktion weiterer Faktoren zur Erhöhung der Varianzaufklärung beitragen würde. Aus diesem Grund wird im Anschluss an die Beschreibung der einfaktoriellen Lösung auch die Möglichkeit einer mehrfaktoriellen Lösung diskutiert werden. (Backhaus/Erichson/Plinke/Weiber (2006), S. 357; Bühner (2011), S. 338; Mummendey/Grau (2014), S. 131)

Je nach Item trägt der einzelne Faktor allerdings in unterschiedlich starkem Ausmaß zur Varianzaufklärung bei ($h^2 = .128-.448$). Dabei liegt keine der Kommunalitäten unterhalb von $h^2 = .10$ bzw. der Anteil aufgeklärter Varianz unterschreitet für keines der Items 10%. Es kann somit davon ausgegangen werden kann, dass die Varianz in den Items ausreichend gut durch den einzelnen Faktor aufgeklärt wird. Gleichzeitig ist aber die *Uniqueness*⁵² der Items mit Werten von .552-.872 vergleichsweise hoch ausgeprägt. Ein großer Teil der Varianz in den Items wird somit nicht von dem extrahierten Faktor aufgeklärt, sondern ergibt sich aus systematischen (Spezifität des Items) und unsystematischen (Mess-) Fehlern. (Bühner (2011), S. 358)

Ein weiterer Hinweis auf einen eher geringen Modellfit der einfaktoriellen Lösung ergibt sich aus dem prozentualen Anteil nicht redundanter Residuen. Diese Residuen repräsentieren die Differenz zwischen den beobachteten Itemkorrelationen und den auf Basis der Faktorstruktur geschätzten Itemkorrelationen. Ein hoher Anteil an nicht redundanten Residuen gibt dabei an, „dass die Ladungen in der Faktorenanalyse die beobachteten empirischen Korrelationen nicht gut reproduzieren bzw. abbilden“ (Bühner (2011), S. 367). Mit anderen Worten: Viele nicht redundante Residuen deuten darauf hin, dass das ermittelte Faktorenmodell die Beziehungen unter den in die Berechnung eingehenden Items nicht gut wiedergibt. Für das hier diskutierte

⁵² Vgl. Bühner (2011), S. 300

einfaktorielle Modell ergeben sich insgesamt 40 (43%) nicht redundante Residuen mit absoluten Werten größer 0,05, was auf eine weniger gute Passung der Faktorstruktur hindeutet. (Backhaus/Erichson/Plinke/Weiber (2006), S. 351f; Bühner (2011), S. 366f)

Die eher niedrige Varianzaufklärung durch das eindimensionale Modell ist möglicherweise damit zu begründen, dass die tatsächliche Messintention der Skala durch eine geringe Augenscheinvalidität verschleiert wurde. Hinzu kommt, dass sich die ursprüngliche Skala aus insgesamt 42 Items zusammensetzte, die den Befragten als einzelne Fragebatterie und mit einer Coverstory als Instruktion präsentiert wurde. Es handelt sich in diesem Sinne auch um einen Blindtest. Im Zusammenschluss könnten diese verschiedenen Aspekte bewirken, dass sich bereits im Antwortverhalten des Einzelnen über die gesamte Skala eine erhebliche Varianz einstellt und somit bei ähnlichen Statements nicht immer konsistent geantwortet wurde. Dies lässt vermuten, dass insbesondere der Messfehleranteil der nicht durch den Faktor aufgeklärten Varianz hoch ausfallen sollte, während die Spezifität der Items eher schwach ausgeprägt ist.

Tab. 6.25: Kommunalitäten, *Uniqueness* und Faktorladungen

Item	h^2 ^a		$1-h^2$ ^b	λ^c
	Anfänglich	Extraktion		
v2.6	,286	,236	0,764	,486
v2.7	,269	,216	0,784	,465
v2.10	,170	,136	0,864	,369
v2.11	,201	,156	0,844	,395
v2.13	,235	,218	0,782	,467
v2.14	,200	,160	0,840	,400
v2.15	,314	,287	0,713	,536
v2.17	,158	,128	0,872	,358
v2.23	,244	,233	0,767	,483
v2.24	,205	,170	0,830	,413
v2.32	,330	,384	0,616	,620
v2.37	,247	,201	0,799	,448
v2.41	,269	,282	0,718	,531
v2.42	,374	,448	0,552	,669

^a Kommunalität bzw. Anteil der durch den Faktor erklärten Itemvarianz (Bühner (2011), S. 310)

^b *Uniqueness* bzw. Einzigartigkeit; Varianzanteil, den das Item mit keinem der anderen Items teilt (Bühner (2011), S. 300)

^c Faktorladung

(Quelle: eigene Berechnungen)

Ein anders gelagertes Bild zeigt sich für die Faktorladungen. Diese geben an, „um wie viele Standardabweichungen sich im Durchschnitt die Ausprägung auf dem Item ändert, wenn sich die Ausprägung auf dem Faktor im Durchschnitt um eine Standardabweichung ändert“ (Bühner (2011), S. 300). Da hier nur ein einzelner Faktor extrahiert wurde und deshalb keine weiteren mit diesem Faktor korrelierten Faktoren existieren, können die Faktorladungen hier aber auch als Korrelationen der Items mit dem Faktor interpretiert werden. (Moosbrugger/Kelava (2012), S. 328f; Bühner (2011), S. 301) Als Schwellenwert für die Berücksichtigung eines Items bei der Interpretation eines Faktors wird $\lambda = .30$ angenommen. Keine der Ladungen unterschreitet diesen Schwellenwert, wobei Item 2.42 mit $\lambda = .669$ das Marker-Item des Faktors „Generalisiertes Vertrauen“ darstellt. Insgesamt werden Faktorladungen von $\lambda = .358-.669$ erreicht und es ergeben sich damit mittlere bis starke Zusammenhänge zwischen dem Faktor und den Items. (Bühner (2011), S. 369; Gorsuch (1983); Kline (1979); Klopp (2013), S. 7)

Nach GUADAGNOLI/VELICER ergibt sich eine stabile Faktorstruktur bereits dann, wenn auf die Faktoren 10 bis 12 Variablen mit $\lambda \geq .40$ laden. Ausgehend von diesem Kriterium kann die hier vorliegende Lösung generalisierend interpretiert werden. Nach der von ihnen entwickelten Formel zur Bestimmung der Stabilität der Faktorstruktur wird für die vorliegende Stichprobe

außerdem ein Wert von $FS \approx .90$ ⁵³ ermittelt, was eine gute Übereinstimmung zwischen „wahrer“ und stichprobenbedingter Faktorstruktur anzeigt. (Bortz (2005), S. 523f; Guadagnoli/Velicer (1988), S. 271ff; Klopp (2013), S. 8)

Ausgehend von den hier dargestellten Analyseschritten kann zwar mit einer gewissen Wahrscheinlichkeit darauf geschlossen werden, dass die in der Objektiv-Skala zusammengefassten Items ein einzelnes Konstrukt abbilden und die Skala dementsprechend Eindimensionalität aufweist. Allerdings lassen die geringe Varianzaufklärung, der hohe Anteil systematischer und unsystematischer Fehler sowie der verhältnismäßig hohe Anteil nicht redundanter Residuen auch Zweifel an der einfaktoriellen Lösung aufkommen.

6.2.3.4 Multifaktorielle Lösung

Im Zuge der mehrfaktoriellen Lösung kommt die oblique Rotationstechnik „Promax“ (Kappa-Wert = 4) zum Einsatz, da davon ausgegangen wurde, dass die einzelnen extrahierten Faktoren nicht voneinander unabhängig sind. Hintergrund dieser Überlegung ist die Annahme, dass die resultierenden Faktoren lediglich einzelne Facetten einer gemeinsamen Zieldimension „Generalisiertes Vertrauen“ darstellen und demnach alle in die Analyse eingehenden Items nicht nur mit einem Faktor, sondern auch mit restlichen Faktoren substantiell korrelieren. Ziel der Rotation ist die Herbeiführung einer Einfachstruktur, wobei die einzelnen Items nur auf jeweils einen der extrahierten Faktoren hochladen (Primärladung) und auf allen anderen Faktoren nur niedrige Ladungen (Sekundärladung) aufweisen. (Backhaus/Erichson/Plinke/Weiber (2006), S. 357f; Bortz (2005), S. 547; Bühner (2011), S. 336ff; Klopp (2013), S. 6f; Moosbrugger/Kelava (2012), S. 332)

Vor dem Hintergrund, dass nur sinnvoll zu interpretierende Faktoren extrahiert werden sollten, und der Empfehlung des Kaiser-Guttman-Kriteriums, wurden insgesamt vier Komponenten extrahiert und interpretiert. (Bühner (2011), S. 349)

Die extrahierten Faktoren erklären vor der Rotation ca. 34,8% der Varianz in den Items (Faktor 1: ca. 24,0%, Faktor 2: ca. 4,4%, Faktor 3: ca. 3,5%, Faktor 4: ca. 2,8%), was im Vergleich zum eindimensionalen Faktormodell eine Steigerung um ca. 10% bedeutet. Allerdings ist zu beachten, dass im Falle von obliquen Rotationen die angegebene aufgeklärte Varianz der Faktoren an der Gesamtvarianz der Items die tatsächlich aufgeklärte Varianz der Faktoren an der Gesamtvarianz der Items überschätzt. Dies hat seinen Grund darin, dass aufgrund der Annahme von korrelierten Faktoren jeder Eigenwert einen Teil Varianz enthält, der auch bereits in einem anderen Faktor enthalten ist. Der Anteil aufgeklärter Varianz nach der Rotation ist deshalb in der Regel höher als vor der Rotation (Faktor 1: ca. 18,4%, Faktor 2: ca. 17,6%, Faktor 3: ca. 14,3%, Faktor 4: ca. 14,9% → aufgeklärte Varianz insgesamt: ca. 65,2%)⁵⁴. BÜHNER ist daher der Auffassung, dass bei obliquen Rotation anhand des Eigenwertes lediglich eine Aussage über die „Wichtigkeit“ der Faktoren gemacht werden kann, die Angabe der aufgeklärten Varianz aber wenig sinnvoll sei. (Bühner (2011), S. 358)

Je nach Item trägt die vier Komponenten umfassende Faktorstruktur unterschiedlich stark zur Varianzaufklärung bei ($h^2 = .164-.552$). Dabei unterschreitet der Anteil aufgeklärter Varianz für

⁵³ $FS = 1 - (1,10 * x_1 - 0,12 * x_2 + 0,066)$, mit $x_1 = \frac{1}{\sqrt{n}}$

und $x_2 = \text{minimaler Ladungswert, der bei der Interpretation der Faktoren berücksichtigt wird}$
 → $FS = 1 - \left(1,10 * \frac{1}{\sqrt{220}} - 0,12 * 0,3 + 0,066\right) \approx 0,90$ (Bortz (2005), S. 523f)

⁵⁴ Für Berechnung vgl. Bühner (2011), S. 358

6. Pilotstudie

keines der Items 10%. Die Varianzaufklärung durch die extrahierten Faktoren ist somit ausreichend gut und erreicht ein leicht höheres Niveau als im Falle der einfaktoriellen Lösung. Die *Uniqueness*⁵⁵ der Items rangiert mit Werten von .448-.836 auf einem ähnlich hohen Niveau wie bei der einfaktoriellen Lösung, hat sich aber gleichzeitig leicht verringert. Der Anteil nicht erklärter Varianz (Spezifität und Messfehler) ist somit immer noch hoch, wobei die mehrdimensionale Faktorenlösung verglichen mit der eindimensionalen Faktorenlösung durchaus ein verbessertes Erklärungsmodell für die Beziehungen unter den einzelnen Items bereitstellt. (Bühner (2011), S. 358)

Dieser Eindruck setzt sich bei Betrachtung des Anteils nicht redundanter Residuen fort. Die vierdimensionale Lösung weist lediglich 8 nicht redundante Residuen mit absoluten Werten größer 0,05 auf, was einem Anteil von nur 8% entspricht. Der Modellfit hat sich somit verglichen mit der einfaktoriellen Lösung deutlich verbessert. Die vier extrahierten Komponenten reproduzieren die Beziehung zwischen den Items gut bzw. die Differenz zwischen der auf Basis der Faktorladungen ermittelten Korrelationsmatrix und der ursprünglichen Korrelationsmatrix ist eher gering. (Backhaus/Erichson/Plinke/Weiber (2006), S. 351f; Bühner (2011), S. 366f)

Auch im Rahmen des vier Komponenten umfassenden Faktormodells ist die ausgeprägte *Uniqueness* der Items mit großer Wahrscheinlichkeit auf die Designvorgaben für Objektive Tests, d.h. die Verschleierung der Messintention, im Zusammenspiel mit der Menge der Items zurückzuführen. Insgesamt hat sich der Modellfit durch die Extraktion mehrerer Faktoren allerdings verbessert.

Tab. 6.26: Kommunalitäten, *Uniqueness* und Faktorladungen

Item	h^2 ^a		1- h^2 ^b	λ^c			
	Anfänglich	Extraktion		Faktor 1: <i>Vertrauensbereitschaft</i>	Faktor 2: <i>Misanthropie</i>	Faktor 3: <i>Fairness</i>	Faktor 4: <i>Hilfsbereitschaft</i>
v2.6	,286	,552	,448	,874	-,177	,064	-,154
v2.7	,269	,334	,666	,520	,166	-,157	,012
v2.10	,170	,164	,836	,003	,304	,179	-,022
v2.11	,201	,314	,686	-,086	-,033	,521	,177
v2.13	,235	,324	,676	,193	-,084	,463	,063
v2.14	,200	,384	,616	-,146	-,019	,085	,663
v2.15	,314	,537	,463	-,041	,783	-,032	-,029
v2.17	,158	,166	,834	-,073	,114	,273	,169
v2.23	,244	,289	,711	,150	,237	-,090	,300
v2.24	,205	,272	,728	,075	-,142	,252	,388
v2.32	,330	,405	,595	,260	,295	-,011	,208
v2.37	,247	,402	,598	-,072	,400	,493	-,208
v2.41	,269	,295	,705	,342	,224	,065	,003
v2.42	,374	,434	,566	,315	,241	,146	,112

^a Kommunalität bzw. Anteil der durch den Faktor erklärten Itemvarianz (Bühner (2011), S. 310)

^b *Uniqueness* bzw. Einzigartigkeit; Varianzanteil, den das Item mit keinem der anderen Items teilt (Bühner (2011), S. 300)

^c Faktorladung

(Quelle: eigene Berechnungen)

Ergebnis der Faktorextraktion sollte eine Einfachstruktur sein. Das bedeutet: Ein Item sollte nach Möglichkeit nur auf einem Faktor hoch laden, während die Ladungen auf den übrigen Faktoren niedrig ausfallen. (Moosbrugger/Kelava (2012), S. 332; Bühner (2011), S. 336, S. 350) Um dies zu gewährleisten, erfolgte die Zuweisung der Items zu einem Faktor entlang zweier Kriterien: Die Faktorladung eines Items sollte $\lambda = .30$ nicht unterschreiten bzw. Items wurden

⁵⁵ Vgl. Bühner (2011), S. 300

dem Faktor zugeordnet, auf dem sie die höchste Ladung aufwiesen. Im Zweifel gab die inhaltliche Passung des Items den Ausschlag für die Zuordnung zu einem Faktor. (Bühner (2006), S. 208f)

Die Extraktion ergab vier Faktoren, denen jeweils 3 bis 4 Items zugeordnet wurden. Dabei konnte weitestgehend eine Einfachstruktur realisiert werden. Lediglich die Items v2.23, v2.32 und v2.37 wiesen betragsähnliche Doppelladungen auf. Im Wesentlichen erfolgte die Faktorzuordnung bei diesen Items aufgrund des Kriteriums „höchste Ladung“. Im Falle von Item v2.23 fiel die Entscheidung eindeutig aus, da die höchste Faktorladung gleichzeitig dem Kriterium „ $\lambda \geq .30$ “ entsprach, während die übrigen Ladungen unterhalb dieser Schwelle lagen. Item v2.37 wies dagegen auf zwei Faktoren Ladungen von $\lambda \geq .30$ auf. Hier wurde der Faktor mit der im Vergleich höheren Ladung ausgewählt, was sich auch mit inhaltlichen Erwägungen deckte. Item v2.32 erreichte auf zwei Faktoren Ladungen, die im Vergleich die geringste Differenz zueinander besitzen, dabei aber in beiden Fällen den Schwellenwert von $\lambda = .30$ unterschritten. Das Item wurde deshalb dem Faktor zugeordnet, auf dem es die höhere Ladung aufwies. Auch diese Entscheidung erwies sich als inhaltlich plausibel.

Einen Sonderfall stellt Item v2.17 dar, dass ebenfalls wie Item v2.32 auf keinem der Faktoren eine Ladung von $\lambda \geq .30$ aufweist. Auch hier wurde entsprechend der inhaltlichen Passung und aufgrund der im Vergleich höchsten Faktorladung über die Zuweisung zu einem Faktor entschieden.

Die Ladungen der Faktor 1: *Vertrauensbereitschaft* zugeordneten Items erreichen Werte von $\lambda = .315$ -.874, wobei Item v2.6 das Marker-Item darstellt. Items, die auf Faktor 2: *Misanthropie* hoch laden, nehmen Ladungswerte von $\lambda = .295$ -.783 an. Als Marker-Item dieses Faktors kann das Item v2.15 bezeichnet werden. Die Ladungen, der auf Faktor 3: *Fairness* hoch ladenden Items nehmen Werte von $\lambda = .273$ -.521 an. Marker-Item ist in diesem Fall v2.11. Die höchsten Ladungen auf Faktor 4: *Hilfsbereitschaft* bewegen sich zwischen $\lambda = .300$ und $\lambda = .663$. Das Marker-Item stellt hierbei v2.14 dar. Es ergibt sich nun die folgende Item-Faktor-Zuordnung:

Tab. 6.27: Item-Faktor-Zuordnung (Objektiv-Skala)

Faktor 1: <i>Vertrauensbereitschaft</i>	Faktor 2: <i>Misanthropie</i>	Faktor 3: <i>Fairness</i>	Faktor 4: <i>Hilfsbereitschaft</i>
2.6) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen.	2.15) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.	2.11) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.	2.14) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.
2.7) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken.	2.10) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinandertreffen, ist von andauernden Konflikten geprägt.	2.37) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.	2.24) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist.
2.41) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.	2.32) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.	2.13) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.	2.23) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden.
2.42) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.	---	2.17) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte.	---

Hierbei ist allerdings zu bemerken, dass das Ergebnis vor dem Hintergrund der bereits weiter oben beschriebenen Kriterien von GUADAGNOLI/VELICER nicht als generalisierend interpretiert werden kann. (Guadagnoli/Velicer (1988), S. 271ff; Klopp (2013), S. 8)

Wie aus der nachfolgenden Korrelationsmatrix sichtbar wird, weisen die einzelnen extrahierten Faktoren mittlere bis starke Zusammenhänge auf. Es soll daher in einem nächsten Analyseschritt überprüft werden, ob eine erneute Faktorisierung der extrahierten Faktoren einen generellen übergeordneten Faktor zu Tage bringt.

Tab. 6.28: Korrelationsmatrix: Faktoren

Faktor	1	2	3	4
1	1	---	---	---
2	,580	1	---	---
3	,454	,439	1	---
4	,556	,521	,424	1

6.2.3.4.1 Prüfung auf generellen übergeordneten Faktor

Grundannahme obliquen Rotationstechniken ist es, dass die extrahierten Faktoren miteinander zusammenhängen bzw. nicht unabhängig voneinander sind. Ergeben sich ausreichend hohe Korrelationen zwischen diesen Faktoren erster Ordnung, so ist denkbar, dass diese Zusammenhänge auf allgemeinere abstraktere Faktoren höherer Ordnung zurückzuführen sind. Im vorliegenden Fall kann daher angenommen werden, dass der Zusammenhang zwischen den

6. Pilotstudie

Faktoren erster Ordnung durch einen gemeinsamen Faktor zweiter Ordnung bedingt ist. Der Faktor zweiter Ordnung ergibt sich dabei infolge der Analyse der Korrelationen zwischen den Faktoren erster Ordnung. (Revenstorf (1976), S. 222, S. 297; Überla (1977), S. 227ff)

Wie bereits bei der Ermittlung der Faktoren erster Ordnung kommt auch bei der Ermittlung der Faktoren zweiter Ordnung eine Hauptachsenanalyse zum Einsatz. Eingang in die Hauptachsenanalyse finden dabei jeweils die Summenscores der vier Faktoren erster Ordnung, sodass insgesamt vier Variablen (Vertrauensbereitschaft/Vertrauenswürdigkeit; Misanthropie; Fairness; Hilfsbereitschaft) in die faktorenanalytische Untersuchung eingehen. Alternativ hätten an dieser Stelle auch die Faktorwerte der Faktoren erster Ordnung verwendet werden können, die die jeweilige Ausprägung einer Person auf einem Faktor bezeichnen und sich vom Summenscore im Wesentlichen durch ihre z-Standardisierung unterscheiden. (Bühner (2011), S. 299)

Wie bereits in den zuvor durchgeführten Hauptachsenanalysen, erfolgt auch die Extraktion der Faktoren zweiter Ordnung auf Basis eines Samples von $N=220$, was nach COMREY/LEE als *fair* bis *good* bewertet werden kann. Auch das Verhältnis von Variablen und Fällen erfüllt mit 4 Variablen zu 220 Fällen (1:55) das von BRYANT/YARNOLD aufgestellten Kriterium zur Stichprobeneignung. (Bryant/Yarnold (1994), S. 100) Die Kriterien von MACCALLUM/WIDAMAN/ZHANG/HONG werden dagegen nicht erfüllt. So erreichen die Kommunalitäten lediglich Werte von $h^2 = .292-.424$ und liegen somit durchweg unterhalb von $h^2 = .50$. (MacCallum/Widaman/Zhang/Hong (1999), S. 96) Es ergibt sich somit nur eine geringe Reliabilität, die auch nicht durch eine Überbestimmtheit der extrahierten Faktoren zweiter Ordnung (hier: 4:1) ausgeglichen werden kann. (MacCallum/Widaman/Zhang/Hong (1999), S. 96) Nach MUNDFROM/SHAW/KE gilt bei niedrigen Kommunalitäten ($h^2 = .20-.40$) und 5 Items pro Faktor bereits eine Fallzahl von $N = 200$ als „gut“. In der hier durchgeführten Analyse kommen 4 Variablen auf einen Faktor bei 220 Fällen. Das Kriterium ist somit nur leicht unterschritten. (Mundfrom/Shaw/Ke (2005), S. 163ff) Insgesamt scheint die in die Analyse eingehende Fallzahl noch ausreichend zu sein.

Tab. 6.29: Maße zur Bestimmung der Stichprobeneignung für EFA^a

N ^b	KMO ^c	MSA ^d	Bartlett-Test ^e		Anti-Image ^f
			Ungefähres χ^2		
220	,729	> .70		139,466	80% ungleich Null
			df	6	
			Signifikanz	,000**	

^a Explorative Faktorenanalyse

^b Fallanzahl

^c Maß der Stichprobeneignung nach KAISER-MEYER-OLKIN (Backhaus/Erichson/Plinke/Weiber (2006), S. 336; Bühner (2011), S. 346f)

^d Wertebereich der MSA-Koeffizienten (*Measure of Sample Adequacy*) (Backhaus/Erichson/Plinke/Weiber (2006), S. 336; Bühner (2011), S. 347, S. 356; Klopp (2013), S. 3f)

^e Bartlett-Test auf Sphärizität (Backhaus/Erichson/Plinke/Weiber (2006), S. 335; Bühner (2011), S. 347f)

^f Anteil der Nicht-diagonalen-Elemente ungleich Null (d.h. > .09) in der Anti-Image-Kovarianzmatrix (Backhaus/Erichson/Plinke/Weiber (2006), S. 335f; Bühner (2011), S. 356f)

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

(Quelle: eigene Berechnungen)

Das KMO zeigt mit einem Wert von ,729 eine mittlere Eignung der eingehenden Variablen für die Faktorenanalyse an. Der MSA-Koeffizient bestätigt diesen Eindruck mit einem Wert, der in den Bereich „middling“ („ziemlich gut“) eingeordnet werden kann. Insgesamt scheinen somit alle in die Analyse eingehenden Variablen für eine faktorenanalytische Untersuchung geeignet zu sein. (Backhaus/Erichson/Plinke/Weiber (2006), S. 336f; Bühner (2011), S. 346f, S. 355; Kaiser/Rice (1974), S. 111ff; Klopp (2013), S. 3f) Aus dem Bartlett-Test auf Sphärizität resultiert

ein hochsignifikantes Ergebnis, was anzeigt, dass die Korrelationen der in die Analyse eingehenden Variablen überzufällig von null abweichen, und unterstreicht damit den auf Basis von KMO und MSA-Koeffizienten gewonnenen Eindruck. (Backhaus/Erichson/Plinke/Weiber (2006), S. 335; Bühner (2011), S. 347f, S. 355)

Einzig das Anti-Image der Varianzen lässt vermuten, dass der Zusammenhang zwischen den Faktoren erster Ordnung nicht auf einen gemeinsamen übergeordneten Faktor zurückzuführen ist. So sind ca. 80% der Nicht-diagonalen Elemente ungleich null, was eine deutliche Überschreitung des von DZIUBAN/SHIRKEY aufgestellten Grenzwertes von 25% bedeutet. (Backhaus/Erichson/Plinke/Weiber (2006), S. 335f; Bühner (2011), S. 356f; Dziuban/Shirkey (1974), S. 359)

Insgesamt sprechen allerdings die Mehrzahl der hier zu Rate gezogen Kriterien für die Opportunität einer Hauptachsenanalyse.

6.2.3.4.1.1 Extraktions- bzw. Abbruchkriterien

Jedes der im Zuge der Faktorisierung der Faktoren erster Ordnung herangezogenen Extraktionskriterien postuliert einen einzigen gemeinsamen Faktor zweiter Ordnung.

Tab. 6.30: Extraktionskriterien für Faktoren

Abbruchkriterium	Anzahl der postulierten Komponenten
Theorie ^a	1
MAP-Test ^b	1
Parallelanalyse ^c	1
Scree-Test ^d	1
Kaiser-Guttman-Kriterium ^e	1

(Quelle: eigene Berechnungen)

Folgt man den theoretischen Annahmen, so sollte jeder der Faktoren erster Ordnung eine Subdimension eines übergeordneten generellen Faktors „Generalisiertes Vertrauen“ darstellen. Dieser einzelne Faktor zweiter Ordnung würde demnach mit der bereits weiter oben vorgestellten einfaktoriellen Lösung erster Ordnung korrespondieren bzw. dieser entsprechen. Gleichzeitig würde sich zeigen, dass die hier geprüfte Skala in letzter Instanz Eindimensionalität aufweisen würde.

Das *Kaiser-Guttman-Kriterium* nimmt nur solche Faktoren als bedeutsam an, die einen Eigenwert größer 1 erreichen. In der vorliegenden Analyse erreichte lediglich ein Faktor einen Eigenwert größer 1 und damit eine höhere Varianzaufklärung als die Einzelvariablen. (Backhaus/Erichson/Plinke/Weiber (2006), S. 353; Bühner (2011), S. 321; Moosbrugger/Kelava (2012), S. 329f; Zwick/Velicer (1986), S. 434, S. 439)

Betrachtet man den Scree-Plot, so erkennt man zwischen dem ersten und dem zweiten Faktor einen Sprung in den Eigenwerten. Zudem liegt der „Knick“ im Scree-Plot unterhalb der Schwelle des Eigenwertkriteriums, weshalb nur der erste extrahierte Faktor als inhaltlich relevant angesehen wurde. (Backhaus/Erichson/Plinke/Weiber (2006), S. 353f; Bühner (2011), S. 322f; Cattell (1966); Moosbrugger/Kelava (2012), S. 330; Zwick/Velicer (1986), S. 434)

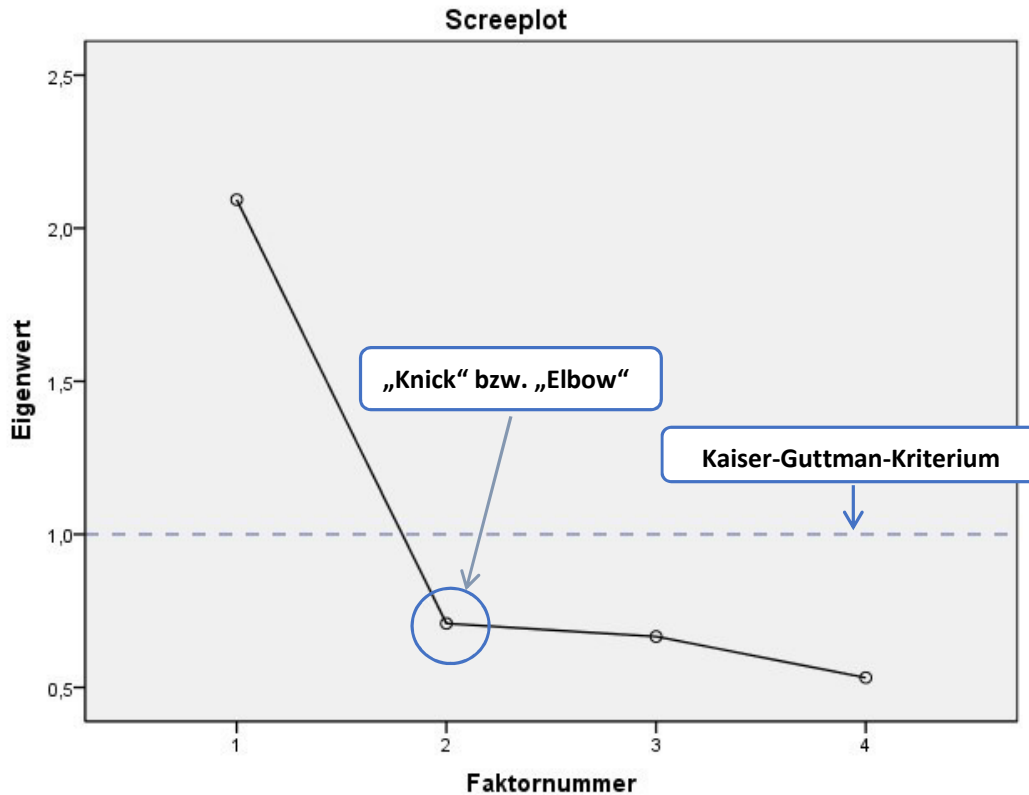


Abb. 6.19: Scree-Test und Kaiser-Guttman-Kriterium (Eigenwert > 1) für Faktorlösung zweiter Ordnung (Objektiv-Skala)

Im Zuge der hier berechneten Parallelanalyse nach HORN wurden insgesamt 1000 Datensets mit je 220 Fällen und 4 Variablen simuliert. Aus den Daten geht hervor, dass lediglich der Eigenwert des ersten Faktors größer als der entsprechende Eigenwert des 95%-Perzentils ist. So erreicht lediglich der empirische Eigenwert des ersten Faktors (*rawdata*) einen höheren Wert als der Verlauf der Eigenwerte, unter dem 95% der Eigenwerte aus Zufallszahlen liegen (*percnytl*), was auf eine einfaktorielle Lösung hindeutet. Dies wird auch durch die zugehörige Grafik gestützt. (Bühner (2011), S. 323ff; Klopp (2013), S. 5f; Moosbrugger/Kelava (2012), S. 331f; O'Connor (2000), S. 397ff; Zwick/Velicer (1986), S. 434f)

Tab. 6.20: Parallelanalyse nach HORN (Ncases = 220; Nvars = 14; Ndatsets = 1000; Percent = 95)

Root ^a	Rawdata ^b	Means ^c	Percnytl ^d
1	2,093235	1,149941	1,238150
2	,708857	1,041647	1,095304
3	,666138	,955467	1,000837
4	,531770	,852945	,923596

^a Faktor (O'Connor (2000); Bühner (2011), S. 324)

^b empirische Eigenwerte (O'Connor (2000); Bühner (2011), S. 324)

^c gemittelte zufallsgenerierte Eigenwerte (O'Connor (2000); Bühner (2011), S. 324)

^d Eigenwerte, unter denen 95% der zufallsgenerierten Eigenwerte liegen (O'Connor (2000); Bühner (2011), S. 324)

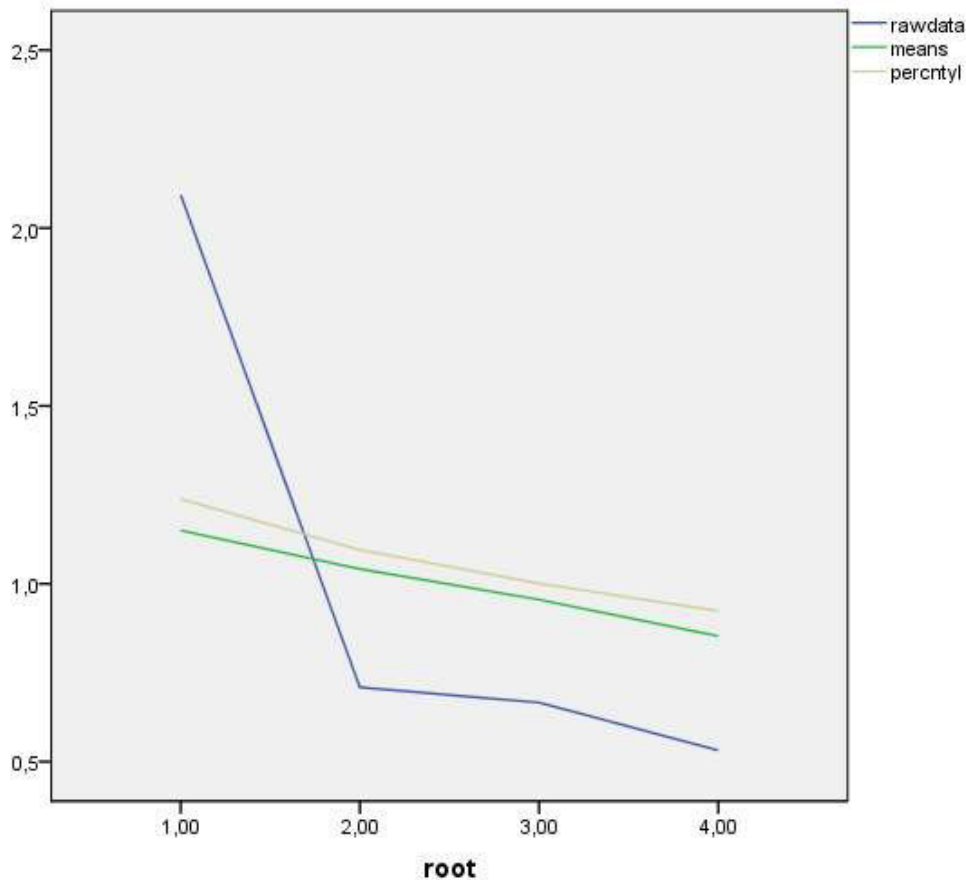


Abb. 6.18: Ergebnis der *Parallelanalyse nach Horn* mit empirisch beobachtetem Eigenwertverlauf (*rawdata*), mittlerem zufallsgeneriertem Eigenwertverlauf (*means*) und 95%-Perzentil der Verteilung zufällig erzeugter Eigenwerte (*percnyl*) für Faktoriellösung zweiter Ordnung (Objektiv-Skala)

Das Ergebnis des *Minimum-Average-Partial-Test* von VELICER zeigt, dass die mittlere quadrierte Partialkorrelation mit der Auspartialisierung des ersten Faktors minimal wird und danach wieder ansteigt. Die systematische Varianz der Korrelationsmatrix wird somit bereits durch einen einzigen Faktor ausgeschöpft. Entsprechend plädiert der MAP-Test für eine einfaktorielle Lösung. (Bühner (2011), S. 325f, S. 363f; O'Connor (2000), S. 397; Zwick/Velicer (1986), S. 434)

Tab. 6.32: MAP-Test von VELICER (*Minimum-Average-Partial-Test*)

Auspartialisierter Faktor	$r^2_{\text{part}}^a$
0	,1342
1	,1154
2	,3678
3	1,0000

^a Quadrierte mittlere Partialkorrelation (Bühner (2011), S. 325f)

Eine Zusammenschau der Abbruchkriterien ergibt, dass die Faktoren erster Ordnung mit großer Wahrscheinlichkeit Subdimensionen einer einzelnen übergeordneten Komponente zweiter Ordnung sind. Eine solche einfaktorielle Lösung höherer Ordnung ist im Speziellen vor dem Hintergrund der theoretischen Annahmen opportun, wird doch davon ausgegangen, dass sämtliche Items der hier geprüften Skala in letzter Instanz eine einzige Dimension, nämlich

„Generalisiertes Vertrauen“ abbilden. Es wird dementsprechend nachfolgend eine eindimensionale Faktoriellösung zweiter Ordnung interpretiert werden.

6.2.3.4.1.2 Einfaktorielle Lösung

Der extrahierte Faktor zweiter Ordnung erklärt durchschnittlich ca. 36,63% der Varianz in den Faktoren erster Ordnung. Je nach Faktor erster Ordnung trägt der Faktor zweiter Ordnung allerdings unterschiedlich stark zur Varianzaufklärung bei ($h^2 = .292-.424$). Dabei unterschreitet der Anteil aufgeklärter Varianz für keinen der Faktoren erster Ordnung die Schwelle von 10%. Es kann somit davon ausgegangen werden, dass die Varianz in den Faktoren erster Ordnung ausreichend gut durch den Faktor zweiter Ordnung aufgeklärt wird. Gleichzeitig ist aber die *Uniqueness*⁵⁶ der Items mit Werten von .576-.708 vergleichsweise hoch ausgeprägt. Ein großer Teil der Varianz in den Faktoren erster Ordnung wird somit nicht von dem extrahierten Faktor aufgeklärt, sondern ergibt sich aus systematischen (Spezifität des Faktors) und unsystematischen (Mess-) Fehlern. (Bühner (2011), S. 358)

Tab. 6.33: Kommunalitäten, *Uniqueness* und Faktorladungen

Faktor erster Ordnung	h^{2a}		$1-h^{2b}$	λ^c
	Anfänglich	Extraktion		
Vertrauensbereitschaft	,270	,424	,576	,651
Misanthropie	,266	,409	,591	,640
Fairness	,224	,339	,661	,583
Hilfsbereitschaft	,199	,292	,708	,541

^a Kommunalität bzw. Anteil der durch den Faktor erklärten Itemvarianz (Bühner (2011), S. 310)

^b *Uniqueness* bzw. Einzigartigkeit; Varianzanteil, den das Item mit keinem der anderen Items teilt (Bühner (2011), S. 300)

^c Faktorladung

(Quelle: eigene Berechnungen)

Ein Hinweis auf einen trotz allem guten Modellfit der einfaktoriellen Lösung ergibt sich aus dem prozentualen Anteil nicht redundanter Residuen. Im vorliegenden Fall ergeben sich 0% nicht redundante Residuen, was anzeigt, dass die auf Basis der Faktorladungen ermittelte Korrelationsmatrix kaum Differenzen zur ursprünglichen Korrelationsmatrix aufweist. Mit anderen Worten: Der Faktor zweiter Ordnung reproduziert den Zusammenhang zwischen den Faktoren erster Ordnung „sehr gut“. (Backhaus/Erichson/Plinke/Weiber (2006), S. 351f; Bühner (2011), S. 366f)

Nach GUADAGNOLI/VELICER ergibt sich eine stabile Faktorstruktur bereits dann, wenn auf jedem bedeutsamen Faktor mindestens 4 Variablen Ladungen von $\lambda \geq .60$ aufweisen. Im vorliegenden Fall erreichen zwei der Faktoren erster Ordnung Ladungen von $\lambda > .60$ auf dem Faktor zweiter Ordnung und die verbleibenden zwei Faktoren erster Ordnung Ladungen von $\lambda > .50$, wobei beide eine Tendenz in Richtung $\lambda = .60$ aufweisen. Ausgehend von diesem Kriterium kann die hier vorliegende Lösung tendenziell als generalisierend interpretiert werden. Nach der von ihnen entwickelten Formel zur Bestimmung der Stabilität der Faktorstruktur wird für die vorliegende Stichprobe außerdem ein Wert von $FS \approx .92$ ⁵⁷ ermittelt, was eine gute

⁵⁶ Vgl. Bühner (2011), S. 300

⁵⁷ $FS = 1 - (1,10 * x_1 - 0,12 * x_2 + 0,066)$, mit $x_1 = \frac{1}{\sqrt{n}}$

und $x_2 = \text{minimaler Ladungswert, der bei der Interpretation der Faktoren berücksichtigt wird}$

$\rightarrow FS = 1 - \left(1,10 * \frac{1}{\sqrt{220}} - 0,12 * 0,5 + 0,066\right) \approx 0,92$ (Bortz (2005), S. 523f)

Übereinstimmung zwischen „wahrer“ und stichprobenbedingter Faktorstruktur anzeigt. (Bortz (2005), S. 523f; Guadagnoli/Velicer (1988), S. 271ff; Klopp (2013), S. 8)

6.2.3.4.1.3 Zusammenfassung

Aus den Ergebnissen der Prüfung auf Eindimensionalität lässt sich ableiten, dass die untersuchten Items in letzter Instanz eine einzige Dimension abbilden, die mit „Generalisiertes Vertrauen“ bezeichnet werden soll. Es wurden dabei zwei Wege beschritten, die jeweils zum gleichen Ergebnis kamen. Vorzuziehen ist hierbei allerdings die multifaktorielle Lösung, welche zunächst 4 Faktoren erster Ordnung zu Tage brachte, aus denen wiederum ein einzelner Faktor zweiter Ordnung extrahiert werden konnte. Die einfaktorielle Lösung, mit einem einzelnen Faktor erster Ordnung, trug im Vergleich weniger zur Varianzaufklärung bei und wies einen weniger guten Modellfit auf.

6.2.4 Itemanalyse (Objektiv-Kurzskala)

Im nachfolgenden Abschnitt soll die Möglichkeit der Ableitung einer Kurzskala zur Erfassung Generalisierten Vertrauens aus der im Rahmen der vorangegangenen Item- und Reliabilitätsanalyse resultierten Objektiv-Skala diskutiert werden.

Als Ansatzpunkt wurden hierzu diejenigen Items herangezogen, die in der Trennschärfenanalyse die höchsten Itemtrennschärfen aufwiesen und von denen dementsprechend erwartet werden kann, dass sie die Zieldimension „Generalisiertes Vertrauen“ besonders gut abbilden können. Die Trennschärfen der fraglichen Items entsprechen darüber hinaus dem von MOOSBRUGGER/KELAVA definierten Intervall für „gute“ Trennschärfen ($r_{i(x-i)} = .40-.70$). (Moosbrugger/Kelava (2012), S. 86)

Tab. 6.34: Itemkennwerte der Rohskala: hohe Trennschärfe (N = 220)

Item	P _i ^a	P _m ^b	σ _i ^c	Sch _i ^d	Ex _i ^e	P _{mVar} ^f	r _{i(x-i)} ^g	Sel ^h	SK ⁱ
v2.13	3,0682	,517	,91656	-,136	-,368	,320	,439**	0,439	0,239
v2.15	3,5682	,642	1,11850	-,390	-,702	,490	,440**	0,459	0,197
v2.32	2,2136	,303	,90397	,500	-,097	,143	,447**	0,486	0,247
v2.11	3,0636	,516	,80837	-,169	-,256	,307	,449**	0,449	0,278
v2.41	2,5091	,377	1,04024	,283	-,596	,210	,450**	0,464	0,216
v2.37	3,7500	,688	,85808	-,457	-,107	,518	,465**	0,502	0,271
v2.42	2,8636	,466	1,02460	,251	-,434	,282	,493**	0,494	0,241

Sortiert nach $r_{i(x-i)}$

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)

^e Exzess bzw. Kurtosis (Ex_i < 0 → breitgipflige Verteilung; Ex_i > 0 → schmalgipflige Verteilung) (Bühner (2006), S. 88f)

^f Mehrstufiger Schwierigkeitsindex (mit Varianzkorrektur) nach FISSENI (Fisseni (1997), S. 45)

^g Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^h Selektionskennwert nach LIENERT/RAATZ (Fisseni (1997), S. 60ff; Lienert/Raatz (1998), S. 118)

ⁱ Selektionskennwert nach BÜHNER (Bühner (2006), S. 99f)

*α<.05 (einseitig), **α<.01 (einseitig)

(Quelle: eigene Berechnungen)

Homogene Tests sollten neben Items mit möglichst hohen Trennschärfen auch Items enthalten, die es erlauben zwischen verschiedenen Merkmalsabstufungen zu differenzieren. Erreicht werden kann dies, wenn die Skala sowohl leichte als auch schwere Items enthält. FISSENI schlägt hierzu ein Intervall von P_m = .40-.60 vor. Drei von den ausgewählten Items weisen

Schwierigkeiten in diesem Bereich auf, die verbleibenden Items erreichen gleichermaßen niedrigere wie höhere Schwierigkeiten. Insgesamt sind die selektierten Items somit in der Lage einen besonders breiten Merkmalsbereich abzudecken. (Fisseni (1997), S. 60)

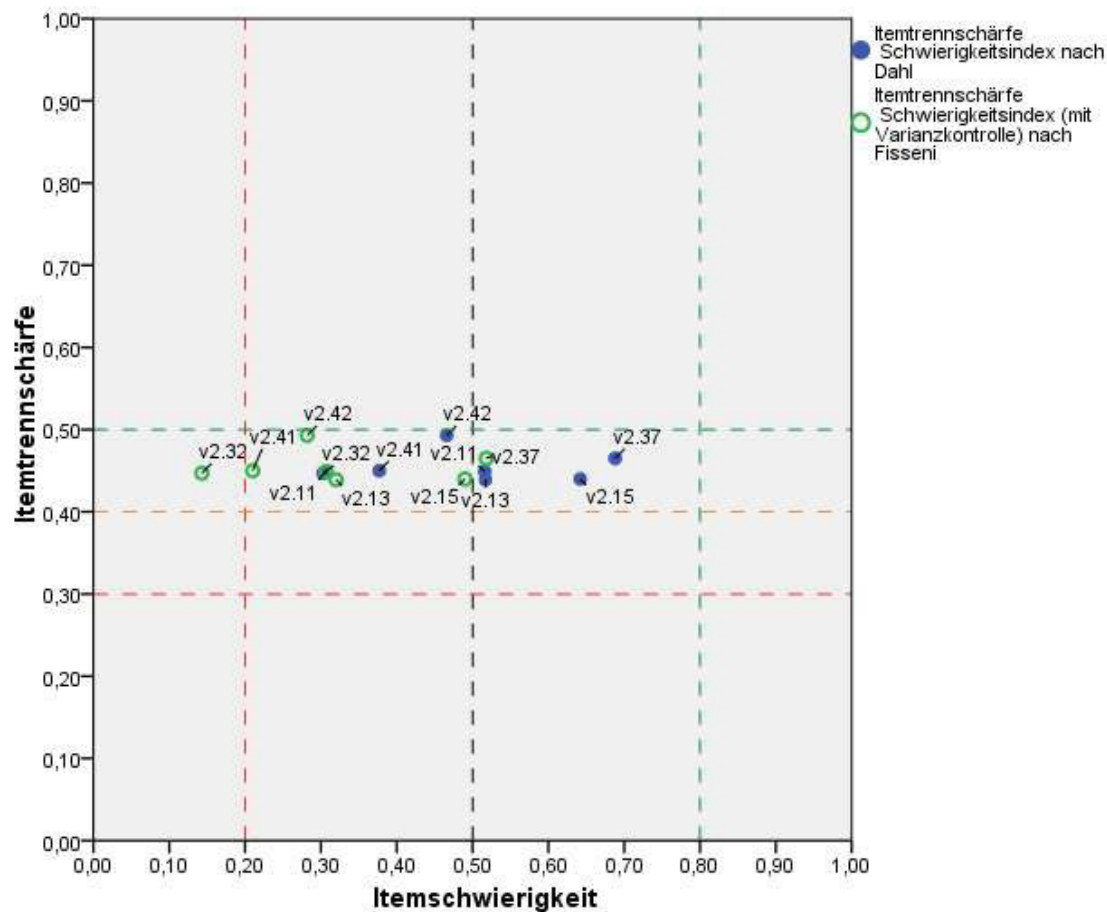


Abb. 6.19: Zusammenhang zwischen Itemschwierigkeit (Schwierigkeitsindex nach DAHL [P_m] bzw. FISSENI [P_{mVar}]) und Itemtrennschärfe (korrigierte Item-Skala-Korrelation [$r_{i(x-i)}$]) für Items der Rohskala „Generalisiertes Vertrauen“ mit hoher, hochsignifikanter Trennschärfe ($r_{i(x-i)} = .40-.70$); ($P > .80$ = Niedrig; $P = .80-.20$ = Mittel; $P < .20$ = Hoch) und ($r_{i(x-i)} < .30$ = Niedrig; $r_{i(x-i)} = .30-.50$ = Mittel; $r_{i(x-i)} > .50$ = Hoch) (Bühner (2011), S. 81)

6.2.5 Reliabilitätsanalyse (Objektiv-Kurzskala)

Wie bereits bei der Objektiv-Skala erfolgt auch hier die Bestimmung der Reliabilität der Objektiv-Kurzskala zunächst anhand der Schätzung der inneren Konsistenz der Skala. Zur Absicherung der Ergebnisse wird außerdem noch die *Split-Half*-Reliabilität bestimmt.

6.2.5.1 Innere Konsistenz

Im Einklang mit der Literatur wird im Rahmen der Bestimmung der inneren Konsistenz $\alpha = .70$ als untere Grenze für den *Cronbach- α* -Koeffizienten angenommen. (Bühner (2011), S. 80; Cortina (1993), S. 101; Schmitt (1996), S. 351; Spector (1992), S. 32)

Im Zuge der Alpha-Maximierung werden Items mit niedrigen Trennschärfen ($r_{i(x-i)} < .30$) schrittweise aus der Skala entfernt, um auf diese Weise *Cronbach- α* sukzessive zu erhöhen. Zusätzlich werden die mittlere Inter-Item-Korrelation (MIC) und die Präzision von α (P_α) als

6. Pilotstudie

Homogenitätsmaße interpretiert. Bei $MIC = .20-.40$ kann ein optimaler Homogenitätslevel für die Skala angenommen werden. Die Präzision von α kann als Standardfehler von *Cronbach- α* interpretiert werden und gilt bei besonders kleinen Werten ($P_\alpha < .01$) als Hinweis auf die Eindimensionalität der Skala. (Briggs/Cheek (1986), S. 114f; Bühner (2006), S. 133; Bühner (2011), S. 168; Cortina (1993), S. 100ff; Moosbrugger/Kelava (2012), S. 133f; Streiner (2003), S. 102)

Eindimensionalität kann demnach angenommen werden, wenn die mittlere Inter-Item-Korrelation angemessen hoch, die Präzision von α sehr niedrig ist, der Streubereich der Item-Korrelationen vollständig im positiven Bereich liegt und die Varianz der Korrelationen niedrig ist. Eine schlussendliche Prüfung auf Eindimensionalität erfolgt im Rahmen einer Faktorenanalyse. (Bühner (2006), S. 133; Bühner (2011), S. 168; Cortina (1993), S. 100ff; Moosbrugger/Kelava (2012), S. 133f)

Tab. 6.35: Inter-Item-Korrelationen und Reliabilitätsstatistiken (N = 220)

	MIC ^a	Min ^b	Max ^b	σ^2_{MIC} ^c	σ_{MIC} ^d	α^e	α_{st} ^f	P_α ^g	n_{Item} ^h
Iteration 0	,277	,152	,416	,006	,0775	,728	,728	,0173	7

^a Mean-Inter-Item-Korrelation bzw. mittlere Inter-Item-Korrelation (Briggs/Cheek (1986), S. 114f)

^b Wertebereich innerhalb dem die Inter-Item-Korrelationen streuen (Briggs/Cheek (1986), S. 114)

^c Varianz der Inter-Item-Korrelationen

^d Standardabweichung der Inter-Item-Korrelationen

^e *Cronbach- α*

^f *Cronbach- α* für standardisierte Items

^g Präzision von *Cronbach- α* (Cortina (1993), S. 100; Bühner (2006), S. 133, S. 145)

^h Anzahl der Items

(Quelle: eigene Berechnung)

Tab. 6.36: Item-Skala-Statistiken (Iteration 0; N = 220)

Iteration 0			
Item	$r_{1(x-j)}^a$	$R^2_{mult}^b$	α_{n-j}^c
v2.11	,345**	,161	,717
v2.13	,367**	,168	,713
v2.15	,452**	,241	,695
v2.32	,491**	,264	,685
v2.37	,448**	,223	,696
v2.41	,443**	,222	,696
v2.42	,543**	,307	,670

^a Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

^b Quadrierte multiple Korrelation

^c *Cronbach- α* , wenn Item gelöscht

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

(Quelle: eigene Berechnung)

Iteration 0 enthält alle 7 Items, die im Rahmen der Trennschärfenanalyse die höchsten Item-Skala-Korrelationen erreichten. Durch den Ausschluss der übrigen 35 Items aus der Ursprungsskala und die Zusammenfassung der genannten Items zur Objektiv-Kurzskala haben sich die Trennschärfen der 7 Items allerdings verändert.

Die Kurzskala erweist sich mit $\alpha = .728$ als ausreichend reliabel. Da sowohl das nicht standardisierte als auch das standardisierte *Cronbach- α* den gleichen Wert annehmen, kann davon ausgegangen werden, dass die Itemvarianzen identisch sind. Betrachtet man die resultierenden Trennschärfen, so lässt sich erkennen, dass jedes der Items eine Item-Skala-Korrelation im mittleren oder hohen Wertebereich aufweist. In der Folge ist keine Alpha-Maximierung notwendig bzw. möglich, da die Entfernung eines Items aus der Skala *Cronbach- α* in jedem Fall verringern würde. (Bühner (2011), S. 241, S. 249)

Mit einer mittleren Inter-Item-Korrelation von $MIC = .277$ ist es zulässig auf eine optimale Homogenität der Skala zu schließen. Außerdem liegt der Streubereich der Item-Korrelationen

(Min = .152; Max = .416) vollständig innerhalb eines positiven Werteintervalls, was einen Hinweis auf die Eindimensionalität der Skala darstellt. Die Streuung der Item-Korrelationen ist mit $\sigma_{MIC} = .0775$ eher gering zu bewerten und zeigt an, dass alle Items wechselseitig etwa gleichstark miteinander zusammenhängen. Die Präzision von α liegt mit $P_\alpha = .0173$ über dem Grenzwert für homogene Tests und kann als Indiz für einen heterogenen Test gedeutet werden. Ausgehend von diesem P_α -Wert besteht somit die Möglichkeit, dass die Objektiv-Kurzskala mehr als nur eine Dimension enthält bzw. abbildet. (Briggs/Cheek (1986), S. 114f; Bühner (2006), S. 133, S. 144f, S. 148; Bühner (2011), S. 243)

Zwar stellt die Überschreitung des Grenzwertes für die Präzision von α einen starken Hinweis auf eine Mehrdimensionalität der Skala dar, gleichzeitig ergeben sich mit den Ausprägungen der übrigen Parameter auch Indizien dafür, dass durch die Items nur eine einzelne Komponente abgebildet wird. Entsprechend kann noch nicht abschließend bewertet werden, ob es sich bei der Kurzskala um eine homogene oder heterogene Skala handelt.

6.2.5.2 Split-Half-Reliabilität

Die Berechnung der Split-Half-Reliabilität erfolgt zur Absicherung der im Rahmen der Schätzung der inneren Konsistenz ermittelten Reliabilität der Skala. Wie bereits beschrieben, wird hierzu die Objektiv-Kurzskala zunächst in zwei möglichst parallele Testhälften unterteilt. Testhälften werden einmal mit Hilfe des Verfahrens der Itemzwillinge und einmal mit Hilfe der Odd-Even-Methode erzeugt. Da die Kurzskala eine ungerade Anzahl von Items enthält, können hierbei keine exakt parallelen Testhälften erzeugt werden. Die zweimalige Schätzung der Split-Half-Reliabilität auf Basis unterschiedlicher Testaufteilungen geschieht, um Abweichungen in der Parallelität der Itemparameter, die sich aus der Zuordnung zu den Testhälften ergeben, auszugleichen. (Bühner (2011), S. 157; Lienert/Ratz (1998), S. 183f; Moosbrugger/Kelava (2012), S. 129f)

Die Reliabilität der Kurzskala wird mit Hilfe eines sogenannten Halbttest-Reliabilitätskoeffizienten geschätzt, der dann in Form des *Spearman-Brown-Koeffizienten* und des *Guttman-Split-Half-Koeffizienten*, auf die Gesamtlänge der Skala aufgewertet wird. Der *Spearman-Brown-Koeffizient* stellt hierbei sowohl Werte für gleich lange Testhälften wie für ungleich lange Testhälften zur Verfügung. Der *Guttman-Split-Half-Koeffizient* stellt dagegen eine konservativere Schätzung der Reliabilität dar und gleicht Abweichungen in der Varianz der Testhälften aus. (Bühner (2011), S. 161ff; S. 236; Kristof (1983), S. 557; Lienert/Ratz (1998), S. 182ff; Moosbrugger/Kelava (2012), S. 128ff)

Zur Bestimmung der Testhälften mit der Odd-Even-Methode wurden den Items zunächst neue Reihungsnummern entsprechend ihrer Reihung in der Ursprungsskala zugeordnet. Danach wurden Items mit ungerader Reihungsnummer in *Testhälfte 1* und Items mit gerader Reihungsnummer in *Testhälfte 2* verschoben. Hieraus resultiert, dass *Testhälfte 1* insgesamt 4 Items und *Testhälfte 2* nur 3 Items enthält:

Tab. 6.37: Testhalbierung: Odd-Even

Testhälfte 1		Testhälfte 2	
Reihungsnummer	Item	Reihungsnummer	Item
1	v2.11	2	v2.13
3	v2.15	4	v2.32
5	v2.37	6	v2.41
7	v2.42	---	---

6. Pilotstudie

Testhälften bestehend aus Itemzwillingen wurden anhand von Itemschwierigkeit, Itemstreuung und Itemtrennschärfe gebildet. Dabei sollten in jeder der beiden Testhälften Items vorhanden sein, für die hinsichtlich der genannten Parameter möglichst exakte Entsprechungen in der anderen Testhälfte existieren. Nach der Bestimmung der Itemzwillinge erfolgte die Zuteilung zu den Testhälften per Zufallsauswahl:

Tab. 6.38: Testhalbierung: Itemzwillinge

Testhälfte 1					Testhälfte 2				
Item	P _i ^a	P _m ^b	σ _i ^c	r _{i (x-ij)} ^d	Item	P _i ^a	P _m ^b	σ _i ^c	r _{i (x-ij)} ^d
v2.11	3,0636	,516	,80837	,345**	v2.13	3,0682	,517	,91656	,367**
v2.15	3,5682	,642	1,11850	,452**	v2.37	3,7500	,688	,85808	,448**
v2.32	2,2136	,303	,90397	,491**	v2.42	2,8636	,466	1,02460	,543**
v2.41	2,5091	,377	1,04024	,443**	---	---	---	---	---

^a Itemschwierigkeit (Mittelwert)

^b Mehrstufiger Schwierigkeitsindex nach DAHL (Dahl (1971); Fisseni (1997), S. 43f)

^c Itemstreuung (Standardabweichung)

^d Korrigierte Itemtrennschärfe (Korrigierte Item-Skala-Korrelation bzw. Part-Whole-Korrelation)

*p<.05 (einseitig), **p<.01 (einseitig)

(Quelle: eigene Berechnungen)

Die nach der Odd-Even-Methode bestimmten Testhälften erreichen eher niedrige Reliabilitäten ($\alpha < .70$), wobei die innere Konsistenz von *Testhälfte 1* höher als bei *Testhälfte 2* ausfällt. Dies ist vermutlich auf die ungleiche Itemzahl innerhalb der beiden Testhälften zurückzuführen. Die Testhälften korrelieren verhältnismäßig stark miteinander. Die Aufwertung der Halb-Test-Reliabilität ergibt $r_{tt} = .721$ bzw. $\rho_{tt} = .705$ und kann jeweils als „akzeptabel“ eingeordnet werden.

Die auf Basis der Itemzwillinge zusammengestellten Testhälften erreichen ebenfalls eher niedrige Reliabilitäten ($\alpha < .70$), wobei auch hier *Testhälfte 1* eine etwas höhere innere Konsistenz aufweist als *Testhälfte 2*. Dies ist wohl ebenfalls der unterschiedlichen Itemanzahl in den Testhälften geschuldet. Die Testhälften korrelieren etwas stärker miteinander als bei der Odd-Even-Methode und auch die aufgewerteten Reliabilitäten liegen mit $r_{tt} = .743$ bzw. $\rho_{tt} = .726$ etwas höher. Beide Werte sind allerdings ebenfalls als „akzeptabel“ zu bewerten.

Tab. 6.39: Reliabilitätsstatistiken

Testhalbierungsmethode	α ₁ ^a	α ₂ ^b	r ₁₂ ^c	r _{tt} ^d	ρ _{tt} ^e
Odd-Even	,616	,527	,560	,721	,705
Itemzwillinge	,585	,545	,587	,743	,726

^a Cronbach-α für Testhälfte 1

^b Cronbach-α für Testhälfte 2

^c Korrelation zwischen den Testhälften

^d Spearman-Brown-Koeffizient (Lienert/Raatz (1998), S. 185; Bühner (2011), S. 162; Moosbrugger/Kelava (2012), S. 128f)

^e Guttman-Split-Half-Koeffizient (Lienert/Raatz (1998), S. 186; Bühner (2011), S. 164)

(Quelle: eigene Berechnungen)

Vergleicht man die Split-Half-Reliabilitäten mit der im Rahmen der Bestimmung der inneren Konsistenz ermittelten Reliabilität, so liegt die konservativere Schätzung mittels des *Guttman-Split-Half-Koeffizienten* immer unterhalb der mittels Cronbach-α ermittelten Reliabilität für die Gesamtskala ($\alpha = .728$). Der *Spearman-Brown-Koeffizient* für ungleich lange Testhälften erreicht bei Anwendung der Odd-Even-Methode ebenfalls einen niedrigeren Wert, während der gleiche Parameter bei Anwendung der Methode der Itemzwillinge etwas über dem Cronbach-α der Gesamtskala liegt. Insgesamt sind die Abweichungen hinsichtlich der

Skalenreliabilität zwischen den verschiedenen Verfahren eher gering, sodass die Reliabilität der Kurzskala mit „akzeptabel“ bewertet werden kann.

Für die Objektiv-Kurzskala resultieren die nachfolgenden Items:

Tab. 6.40: Resultat der Reliabilitätsanalyse für Objektiv-Kurzskala

(1)	2.11) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden. (+)
(2)	2.13) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen. (+)
(3)	2.15) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt. (-)
(4)	2.32) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun. (-)
(5)	2.37) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden. (+)
(6)	2.41) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt. (-)
(7)	2.42) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern. (-)

Hoher Itemscore entspricht starker Ausprägung der Zieldimension, niedriger Itemscore entspricht schwacher Ausprägung der Zieldimension. Es ergeben sich daher die folgenden Codierungen für die Antwortskala:

(+) Itempolung positiv: „Stimme nicht zu“ (0); „Stimme eher zu“ (1); „Stimme teils zu, stimme teils nicht zu“ (2); „Stimme eher zu“ (3); „Stimme zu“ (4)

(-) Itempolung negativ: „Stimme nicht zu“ (4); „Stimme eher zu“ (3); „Stimme teils zu, stimme teils nicht zu“ (2); „Stimme eher zu“ (1); „Stimme zu“ (0)

6.2.6 Prüfung auf Eindimensionalität (Objektiv-Kurzskala)

Nachfolgend soll mit Hilfe einer Hauptachsenanalyse geprüft werden, ob es sich bei der Objektiv-Kurzskala um eine homogene eindimensionale Skala oder eine heterogene mehrdimensionale Skala handelt. (Mummendey/Grau (2014), S. 129ff)

Man kann nie von einer messfehlerfreien Messung ausgehen, was dazu führt, dass die Varianz innerhalb der in die Analyse eingehenden Items niemals vollständig durch ein angenommenes Modell aufgeklärt werden kann. Die Hauptachsenanalyse trägt diesem Umstand Rechnung, indem die mit Hilfe dieses Verfahrens extrahierten Faktoren nur die Varianz der Items erklären, die diese mit den anderen Items im Modell teilen. (Backhaus/Erichson/Plinke/Weiber (2006), S. 350f; Bühner (2011), S. 310ff; Klopp (2013), S. 1f; Moosbrugger/Kelava (2012), S. 328)

6.2.6.1 Stichprobeneignung bzw. Opportunität der faktorenanalytischen Untersuchung

Auch in die Hauptachsenanalyse der Kurzskala gehen N = 220 Fälle ein, was nach COMREY/LEE eine als „fair“ bewertbare und damit für eine explorative Faktorenanalyse geeignete Stichprobengröße darstellt. Nach BRYANT/YARNOLD muss ein Verhältnis von 1:3 bis 1:10 zwischen Variablen und Fällen bestehen, damit eine Stichprobe kompatibel ist. Auch dieses Kriterium wird mit 7 Items auf 220 Fälle (Verhältnis von 1:31) mehr als erfüllt. Weitere Kriterien für die Angemessenheit der Fallzahl stellen die anhand der Kommunalitäten bestimmte Item-Reliabilität und der Grad der Überbestimmtheit der Faktoren dar. Erreichen die Kommunalitäten, wie im vorliegenden Fall, Werte von $h^2 < .50$, so sollten möglichst wenige Faktoren herausgefiltert werden, die jeweils durch möglichst viele Items bestimmt sind (6 oder mehr Items pro Faktor). Die Kommunalitäten der für die Kurzskala berechneten einfaktoriellen

6. Pilotstudie

Lösung erreichen lediglich Werte von $h^2 = .10-.50$, wobei die Mehrzahl (ca. 71%) Werte von $h^2 > .20$ annimmt. Für niedrige Kommunalitäten ($h^2 = .20-.40$) gelten nach MUNDFROM/SHAW/KE bereits Verhältnisse von 8 Items pro Faktor auf 130 Fälle oder 6 Items pro Faktor auf 260 Fälle als „ausgezeichnet“. Außerdem werden bei 8 Items pro Faktor bereits 80 Fälle als „gut“ angesehen, ebenso wie 5 Items pro Faktor bei 200 Fällen. Für die Kurzskala sollen 7 Items auf einen einzelnen Faktor bei 220 Fällen laden. Die Stichprobengröße scheint damit ausreichend groß zu sein für die Durchführung einer explorativen Faktorenanalyse. (Bryant/Yarnold (1994), S. 100; Bühner (2011), S. 344f; Cattell (1978), S. 509; Gorsuch (1983), S. 332; Guilford (1954), S. 533; Kline (1979), S. 38ff; MacCallum/Widaman/Zhang/Hong (1999), S. 96; Mundfrom/Shaw/Ke (2005), S. 163ff; Velicer/Fava (1998), S. 231f)

Das KMO gibt an, ob sich zwischen den Items wesentliche Korrelationen ergeben, die eine Faktorenanalyse rechtfertigen. Je höher der Zusammenhang zwischen den Items dabei ist, desto mehr gemeinsame Varianz weisen die Items auf. Für die Kurzskala ergibt sich für das KMO ein Wert von .808, was eine „gute“ Eignung anzeigt. Dieser Eindruck wird zusätzlich durch die MSA-Koeffizienten bestätigt, die mehrheitlich Werte $> .80$ annehmen. Lediglich für Item v2.11 ergibt sich mit $MSA = .784$ ein leicht niedrigerer als „mittel“ bzw. „ziemlich gut“ zu bezeichnender Wert. Es kann somit davon ausgegangen werden, dass die Items insgesamt stark miteinander korrelieren. (Backhaus/Erichson/Plinke/Weiber (2006), S. 336f; Bühner (2011), S. 346f, S. 355; Kaiser/Rice (1974), S. 111ff; Klopp (2013), S. 3f)

Unterstützt werden die vorgenannten Ergebnisse zusätzlich durch das Resultat des Bartlett-Test auf Sphärizität. Die Prüfung darauf, ob die Korrelationen der Items signifikant von null abweichen, konnte hier positiv beschieden werden. (Backhaus/Erichson/Plinke/Weiber (2006), S. 335; Bühner (2011), S. 347f, S. 355)

Tab. 6.41: Maße zur Bestimmung der Stichprobeneignung für EFA^a

N ^b	KMO ^c	MSA ^d	Bartlett-Test ^e		Anti-Image ^f
			Ungefähres χ^2	243,117	
			df	21	
			Signifikanz	,000**	
220	,808	1 > .70 6 > .80			52% ungleich Null

^a Explorative Faktorenanalyse

^b Fallanzahl

^c Maß der Stichprobeneignung nach KAISER-MEYER-OLKIN (Backhaus/Erichson/Plinke/Weiber (2006), S. 336; Bühner (2011), S. 346f)

^d Wertebereich der MSA-Koeffizienten (*Measure of Sample Adequacy*) (Backhaus/Erichson/Plinke/Weiber (2006), S. 336; Bühner (2011), S. 347, S. 356; Klopp (2013), S. 3f)

^e Bartlett-Test auf Sphärizität (Backhaus/Erichson/Plinke/Weiber (2006), S. 335; Bühner (2011), S. 347f)

^f Anteil der Nicht-diagonalen-Elemente ungleich Null (d.h. $> .09$) in der Anti-Image-Kovarianzmatrix (Backhaus/Erichson/Plinke/Weiber (2006), S. 335f; Bühner (2011), S. 356f)

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

(Quelle: eigene Berechnungen)

Das Anti-Image beschreibt im Wesentlichen den Varianzanteil eines Items, der unabhängig von den übrigen ins Modell aufgenommenen Items ist. Im Rahmen der Faktorenanalyse soll nun der Zusammenhang zwischen den Items durch möglichst wenige Faktoren aufgeklärt werden. Hierzu muss das Anti-Image der Varianzen möglichst klein sein. Überprüft wird diese anhand des Anteils der nicht-diagonalen Elemente der Anti-Image-Kovarianz-Matrix, die ungleich null sind ($> .09$), der einen Schwellenwert von 25% nicht überschreiten sollte. Im vorliegenden Fall nehmen allerdings 52% der nicht-diagonalen Elemente Werte ungleich null an. Ausgehend von diesem Ergebnis kann angenommen werden, dass den in die Faktorenanalyse eingehenden Variablen möglicherweise keine gemeinsamen Faktoren zugrunde liegen.

(Backhaus/Erichson/Plinke/Weiber (2006), S. 335f; Bühner (2011), S. 356f; Dziuban/Shirkey (1974), S. 359)

Für die betrachteten Gütekriterien zur Bestimmung der Eignung der Itemauswahl für eine faktorenanalytische Untersuchung zeigen sich vornehmlich günstige Ergebnisse. Lediglich, das Anti-Image weist einen Wert auf, der gegen die Anwendung einer Faktorenanalyse spricht. Insgesamt scheint die Itemauswahl damit ausreichend gut geeignet zu sein.

6.2.6.2 Extraktions- bzw. Abbruchkriterien

Zur Bestimmung der Anzahl der zu extrahierenden Faktoren wurden verschiedene Kriterien herangezogen. In absteigender Reihenfolge ihrer Relevanzsetzung waren dies die sinnvolle Interpretation der Faktoren auf Basis der theoretischen Vorüberlegungen, der *Minimum-Average-Partial-Test* von VELICER (MAP-Test), die *Parallelanalyse* nach HORN, der *Scree-Test* nach CATTELL und das Kaiser-Guttman- bzw. Eigenwertkriterium. Mit Ausnahme des Kaiser-Guttman-Kriteriums sind die Ergebnisse der Prüfungsverfahren deckungsgleich. Somit entspricht das Resultat für die Kurzskaala qualitativ dem für die Gesamtskala.

Tab. 6.42: Extraktionskriterien für Faktoren

Abbruchkriterium	Anzahl der postulierten Komponenten
Theorie ^a	1 bis 4
MAP-Test ^b	1
Parallelanalyse ^c	1
Scree-Test ^d	1 bis 2
Kaiser-Guttman-Kriterium ^e	2

(Quelle: eigene Berechnungen)

Das Kaiser-Guttman-Kriterium nimmt solche Faktoren als relevant an, die Eigenwerte größer 1 aufweisen. Faktoren, die dieses Kriterium erfüllen, klären in der Regel mehr Varianz im Antwortverhalten auf ein Item auf als dieses Item allein. Anhand des Kaiser-Guttman-Kriteriums wird für die Itemauswahl der Objektiv-Kurzskaala eine Lösung mit 2 Faktoren vorgeschlagen. Vor dem Hintergrund der Ergebnisse der übrigen Abbruchkriterien (jeweils 1 Faktor) und der Tatsache, dass das Eigenwertkriterium dazu tendiert, die wahre Anzahl der Faktoren zu überschätzen, ist eine zweifaktorielle Lösung höchstwahrscheinlich zu verwerfen. (Backhaus/Erichson/Plinke/Weiber (2006), S. 353; Bühner (2011), S. 321; Moosbrugger/Kelava (2012), S. 329f; Zwick/Velicer (1986), S. 434, S. 439)

Der erste „Knick“ ergibt sich innerhalb des Scree-Plots⁵⁸ nach dem ersten Faktor, was für eine einfaktorielle Lösung spricht. Außerdem ist die Differenz der Eigenwerte zwischen dem ersten und zweiten Faktor deutlich größer als die Differenzen zwischen den verbleibenden Faktoren. Für eine mehrfaktorielle Lösung spricht indes, dass auch der zweite Faktor noch einen Eigenwert größer 1 aufweist und somit einen größeren Beitrag zur Varianzaufklärung leistet, als dies ein einzelnes Item vermag. Zusätzlich wird der Kurvenverlauf nach dem zweiten Faktor flacher und eine Interpretation des Faktors nach dem „Knick“ als „inhaltlich relevant“ ist durch methodische Konventionen abgedeckt. Zu beachten ist allerdings, dass die Interpretation des Scree-Plots

⁵⁸ Im Rahmen des Scree-Test wird grafisch überprüft, an welcher Stelle die Eigenwerte einen Sprung machen bzw. wo die Differenz zwischen den Eigenwerten zweier Faktoren am größten ist. Innerhalb des Scree-Plots ergibt sich dann ein charakteristischer „Knick“ (Elbow), hinter dem sich die verbleibenden Eigenwerte asymptotisch der x-Achse nähern. Als inhaltlich relevant werden in der vorliegenden Analyse vornehmlich solche Faktoren angenommen, die vor dem ersten „Knick“ angeordnet sind.

6. Pilotstudie

immer auch ein subjektives Element beinhaltet, sodass der Scree-Test allenfalls als unterstützendes grafisches Entscheidungskriterium für zusätzliche Abbruchkriterien angesehen werden sollte. (Backhaus/Erichson/Plinke/Weiber (2006), S. 353f; Bühner (2011), S. 322f; Cattell (1966); Moosbrugger/Kelava (2012), S. 330; Zwick/Velicer (1986), S. 434)

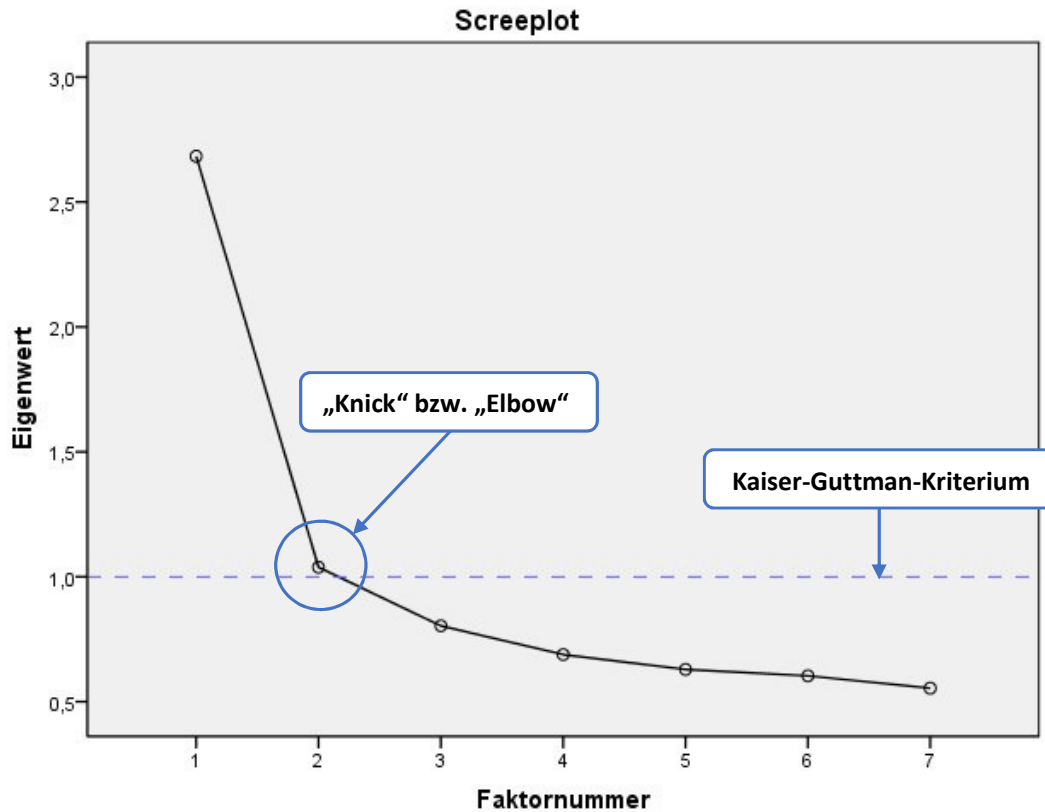


Abb. 6.20: Scree-Test und Kaiser-Guttman-Kriterium (Eigenwert > 1) für Faktorenlösung erster Ordnung (Objektiv-Kurzskala)

Im Rahmen einer Parallelanalyse nach HORN⁵⁹ wurden insgesamt 1000 Datensets mit je 220 Fällen und 7 Variablen simuliert. In der unten aufgetragenen Grafik finden sich der Verlauf der empirischen Eigenwerte (*rawdata*), der Verlauf der gemittelten zufallsgenerierten Eigenwerte (*means*) und der Verlauf der Eigenwerte unter dem 95% der Eigenwerte aus Zufallszahlen liegen (*percntyl*). Ist der empirische Eigenwert größer als dieses 95%-Perzentil, so gehört er zu den extremsten 5% der zufallsgenerierten Eigenwerte und es ist sehr unwahrscheinlich, dass er zufällig zustande gekommen ist. Es handelt sich somit um einen systematischen Faktor. (Bühner (2011), S. 323ff; Klopp (2013), S. 5f; Moosbrugger/Kelava (2012), S. 331f; O'Connor (2000), S. 397ff; Zwick/Velicer (1986), S. 434f)

⁵⁹ Eine Parallelanalyse nach HORN erzeugt eine bestimmte Anzahl von Zufallsdatensätzen aus den Basisdaten und unterzieht diese jeweils einer Faktorenanalyse. In der Folge werden nur diejenigen Faktoren extrahiert werden, deren empirische Eigenwerte größer als die gemittelten Eigenwerte der zufallsgenerierten Daten sind.

Tab. 6.43: Parallelanalyse nach HORN (Ncases = 220; Nvars = 7; Ndatsets = 1000; Percent = 95)

Root ^a	Rawdata ^b	Means ^c	Percntyl ^d
1	2,683075	1,260319	1,356669
2	1,038025	1,149551	1,222052
3	,803750	1,067953	1,120133
4	,688622	,991403	1,038490
5	,629030	,922243	,969345
6	,603396	,848395	,904533
7	,554103	,760137	,829305

^a Faktor (O'Connor (2000); Bühner (2011), S. 324)

^b empirische Eigenwerte (O'Connor (2000); Bühner (2011), S. 324)

^c gemittelte zufallsgenerierte Eigenwerte (O'Connor (2000); Bühner (2011), S. 324)

^d Eigenwerte, unter denen 95% der zufallsgenerierten Eigenwerte liegen (O'Connor (2000); Bühner (2011), S. 324)

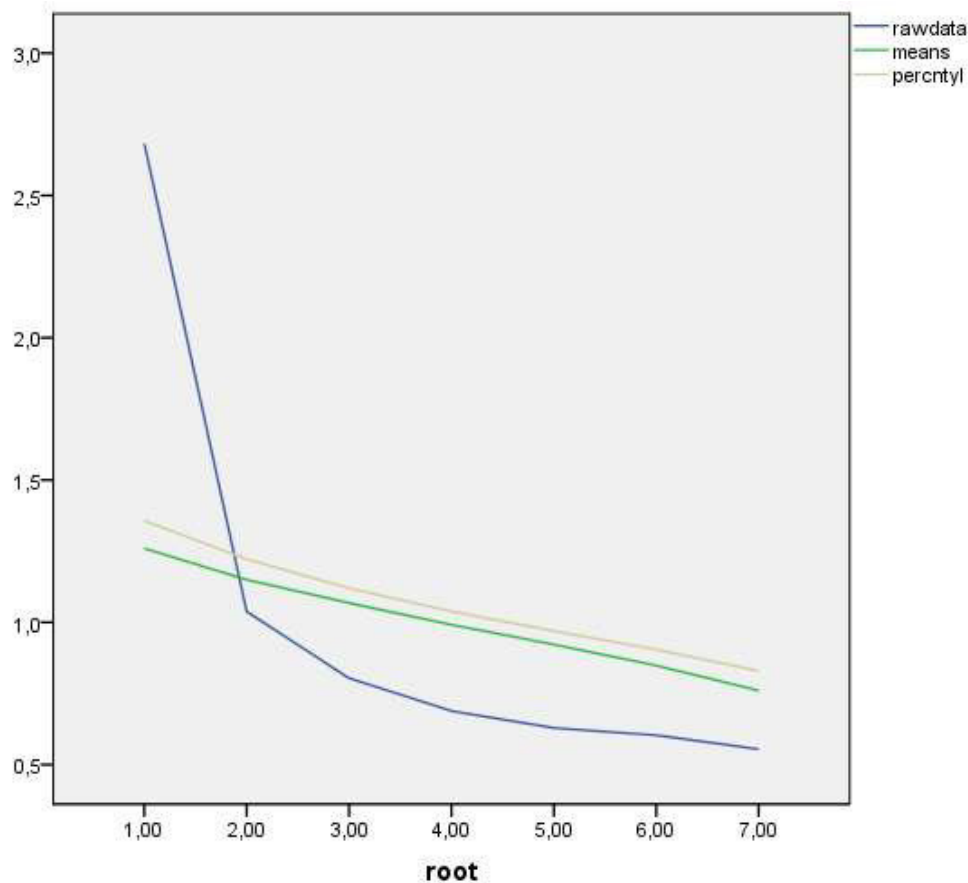


Abb. 6.21: Ergebnis der Parallelanalyse nach Horn mit empirisch beobachtetem Eigenwertverlauf (rawdata), mittlerem zufallsgeneriertem Eigenwertverlauf (means) und 95%-Perzentil der Verteilung zufällig erzeugter Eigenwerte (percnyl) für Faktorenlösung erster Ordnung (Objektiv-Kurzskala)

Aus den Daten geht hervor, dass lediglich der Eigenwert des ersten Faktors größer als der entsprechende Eigenwert des 95%-Perzentils ist. Die Parallelanalyse empfiehlt somit eine einfaktorielle Lösung.

Im Rahmen des *Minimum-Average-Partial-Test* von VELICER wird die Anzahl der zu extrahierenden Faktoren durch die schrittweise Auspartialisierung der mittels einer Hauptkomponentenanalyse ermittelten Faktoren bestimmt. (Bühner (2011), S. 325f, S. 363f; O'Connor (2000), S. 397; Zwick/Velicer (1986), S. 434)

Aus der nachfolgenden Tabelle geht hervor, dass die mittlere quadrierte Partialkorrelation mit der Auspartialisierung des ersten Faktors minimal wird und danach wieder ansteigt. Die systematische Varianz der Korrelationsmatrix wird somit bereits durch einen einzigen Faktor ausgeschöpft. Der MAP-Test empfiehlt dementsprechend eine einfaktorielle Lösung.

Tab. 6.44: MAP-Test von VELICER (*Minimum-Average-Partial-Test*)

Auspartialisierter Faktor	$r^2_{\text{part}}^a$
0	,0825
1	,0348
2	,0722
3	,1333
4	,2501
5	,4835
6	1,0000

^a Quadrierte mittlere Partialkorrelation (Bühner (2011), S. 325f)

Das Abbruchkriterium, welches die größte Relevanz für die Bestimmung der zu extrahierenden Faktorenanzahl besitzt, stellen die theoretischen Vorüberlegungen dar. Da die Objektiv-Kurzskala in der Objektiv-Skala enthalten ist, gilt auch hier die Annahme von Eindimensionalität. (Bühner (2011), S. 321, S. 349; Klopp (2013), S. 4f)

Die Zusammenschau der verschiedenen Abbruchkriterien zeigt, dass die Ergebnisse hinsichtlich der zu extrahierenden Faktorenanzahl für die Parallelanalyse nach HORN und den MAP-Test von VELICER übereinstimmen. Empfohlen wird hiernach eine einfaktorielle Lösung, was auch den Hinweisen aus der zuvor durchgeführten Reliabilitätsanalyse entspricht. Das Kaiser-Guttman-Kriterium weicht mit einer Empfehlung für die Extraktion von 2 Komponenten von den Empfehlungen dieser Abbruchkriterien ab. Gleiches gilt für den Scree-Test, der bis zu zwei Komponenten postuliert. Ausgehend von der theoretischen Operationalisierung müssten innerhalb der Objektiv-Kurzskala bis zu 4 Faktoren enthalten sein. In Anbetracht der Heterogenität der Empfehlungen der Extraktionskriterien sollen nachfolgend sowohl eine einfaktorielle als auch eine multifaktorielle Lösung diskutiert werden. Für eine einfaktorielle Lösung spricht, dass sich eine vergleichsweise starke erste Hauptkomponente ergibt und der durch die übrigen Faktoren bereitgestellte Anteil erklärter Varianz marginal ist. Auf der anderen Seite erhöht eine mehrfaktorielle Lösung den Modellfit und die aufgeklärte Varianz. (Bühner (2011), S. 349)

6.2.6.3 Einfaktorielle Lösung

Da nachfolgend nur ein einzelner Faktor extrahiert wurde, kommt keine Rotationsmethode zum Einsatz. (Bühner (2011), S. 338; Backhaus/Erichson/Plinke/Weiber (2006), S. 357) Der auf diese Weise extrahierte Faktor erklärt durchschnittlich ca. 28,43% der Varianz in den Items. Zum Vergleich wurde eine weitere Hauptachsenanalyse berechnet, bei der den Empfehlungen des Kaiser-Guttman-Kriteriums gefolgt wurde. Es wurde dabei davon ausgegangen, dass die zu extrahierenden 2 Faktoren nicht voneinander unabhängig sind und dementsprechend die oblique Rotationstechnik „Promax“ (Kappa-Wert = 4) angewandt. Ein Vergleich mit der zweifaktoriellen Lösung zeigte, dass die Extraktion weiterer Faktoren eine leichte Erhöhung der Varianzaufklärung bewirken würde. (Backhaus/Erichson/Plinke/Weiber (2006), S. 357; Bühner (2011), S. 338; Mummendey/Grau (2014), S. 131)

Für diese eher gering bis mittlere Varianzaufklärung durch das eindimensionale Modell können im Wesentlichen die gleichen Gründe verantwortlich gemacht werden, wie schon bei der

Analyse der Objektiv-Skala. So scheint die geringe Augenscheinvalidität, die den Vorgaben für Objektive Tests geschuldet ist, und die Breite der Ursprungsskala mit 42 Items für einen erhöhten Messfehleranteil innerhalb der Itemvarianz verantwortlich zu sein.

Tab. 6.45: Kommunalitäten, *Uniqueness* und Faktorladungen

Item	h^{2a}		$1-h^{2b}$	λ^c
	Anfänglich	Extraktion		
v2.11	,161	,163	,837	,403
v2.13	,168	,185	,815	,430
v2.15	,241	,301	,699	,549
v2.32	,264	,345	,655	,588
v2.37	,223	,275	,725	,524
v2.41	,222	,281	,719	,530
v2.42	,307	,440	,560	,664

^a Kommunalität bzw. Anteil der durch den Faktor erklärten Itemvarianz (Bühner (2011), S. 310)

^b *Uniqueness* bzw. Einzigartigkeit; Varianzanteil, den das Item mit keinem der anderen Items teilt (Bühner (2011), S. 300)

^c Faktorladung

(Quelle: eigene Berechnungen)

Ein anderes Bild zeichnen die Faktorladungen. Diese geben an, „um wie viele Standardabweichungen sich im Durchschnitt die Ausprägung auf dem Item ändert, wenn sich die Ausprägung auf dem Faktor im Durchschnitt um eine Standardabweichung ändert“ (Bühner (2011), S. 300). Da hier nur ein einzelner Faktor extrahiert wurde und deshalb keine weiteren mit diesem Faktor korrelierten Faktoren existieren, können die Faktorladungen hier aber auch als Korrelationen der Items mit dem Faktor interpretiert werden. (Bühner (2011), S. 301; Moosbrugger/Kelava (2012), S. 328f) Als Schwellenwert für die Berücksichtigung eines Items bei der Interpretation eines Faktors wird $\lambda = .30$ angenommen. Keine der Ladungen unterschreitet diesen Schwellenwert, wobei Item 2.42 mit $\lambda = .664$, wie schon zuvor bei der Objektiv-Skala, das Marker-Item für die Kurzskaala darstellt. Insgesamt werden Faktorladungen von $\lambda = .403$ -.664 erreicht, was mittlere bis starke Zusammenhänge zwischen dem Faktor und den Items anzeigt. Die Korrelationen zwischen Items und Faktor sind dabei sogar höher als bei der Gesamtskala. (Bühner (2011), S. 369; Gorsuch (1983); Kline (1979); Klopp (2013), S. 7)

Nach GUADAGNOLI/VELICER ergibt sich eine stabile Faktorstruktur, wenn pro Faktor vier Items Ladungen von $\lambda \geq .60$ aufweisen oder auf die Faktoren 10 bis 12 Variablen mit $\lambda \geq .40$ laden. Bei weniger als 10 Variablen mit Ladungen von $\lambda \geq .40$ sollte mindestens eine Stichprobe mit 300 Fällen in die Analyse eingehen, um die Faktorstruktur generalisierend interpretieren zu können. Bei weniger als 10 Variablen, Ladungen von $\lambda < .40$ und weniger als 300 Fällen können zufällige Ladungsmuster auftreten. Da in die Analyse lediglich 7 Items eingehen, kann diesen Kriterien nicht entsprochen werden. Trotz allem gilt nach der von ihnen entwickelten Formel zur Bestimmung der Stabilität der Faktorstruktur für die vorliegende Stichprobe ein Wert von $FS \approx .91^{60}$, was eine gute Übereinstimmung zwischen „wahrer“ und stichprobenbedingter Faktorstruktur anzeigt. Insgesamt bedarf es aber wohl weiterer Untersuchungen mit größeren Stichproben, um von einer Generalisierung der Faktorstruktur der Objektiv-Kurzskaala ausgehen zu können. (Bortz (2005), S. 523f; Guadagnoli/Velicer (1988), S. 271ff; Klopp (2013), S. 8)

Ausgehend von den hier dargestellten Analyseschritten kann mit hoher Wahrscheinlichkeit davon ausgegangen werden, dass die in der Objektiv-Kurzskaala zusammengefassten Items nur

⁶⁰ $FS = 1 - (1,10 * x_1 - 0,12 * x_2 + 0,066)$, mit $x_1 = \frac{1}{\sqrt{n}}$

und $x_2 = \text{minimaler Ladungswert, der bei der Interpretation der Faktoren berücksichtigt wird}$

$\rightarrow FS = 1 - \left(1,10 * \frac{1}{\sqrt{220}} - 0,12 * 0,4 + 0,066\right) \approx 0,91$ (Bortz (2005), S. 523f)

ein einzelnes Konstrukt abbilden und die Skala dementsprechend Eindimensionalität aufweist. Ob es sich bei der gefundenen Faktorstruktur um eine Anordnung handelt, die durch nachfolgende Untersuchungen in gleicher Art und Weise repliziert werden kann, kann an dieser Stelle nicht abschließend beantwortet werden.

6.2.6.4 Multifaktorielle Lösung

Im Rahmen der mehrfaktoriellen Lösung wird die oblique Rotationstechnik „Promax“ (Kappa-Wert = 4) eingesetzt. Es wird somit angenommen, dass die extrahierten Faktoren in einem wechselseitigen Abhängigkeitsverhältnis zueinanderstehen. (Backhaus/Erichson/Plinke/Weiber (2006), S. 357f; Bortz (2005), S. 547; Bühner (2011), S. 336ff; Klopp (2013), S. 6f; Moosbrugger/Kelava (2012), S. 332)

Entsprechend des Kaiser-Guttman-Kriteriums wurden insgesamt zwei Komponenten extrahiert und interpretiert. (Bühner (2011), S. 349) Dabei erklären die gefundenen Faktoren vor der Rotation ca. 34,49% der Varianz in den Items (Faktor 1: ca. 29,09%; Faktor 2: ca. 5,4%), was im Vergleich zum eindimensionalen Faktormodell eine Steigerung um ca. 6% anzeigt. Nach BÜHNER ergeben sich für den Anteil erklärter Varianz nach der Rotation folgende Werte: Faktor 1: ca. 27,3%; Faktor 2: ca. 21,0%. Entsprechend besitzt Faktor 1 eine größere Relevanz. (Bühner (2011), S. 358)

Tab. 6.46: Kommunalitäten, *Uniqueness* und Faktorladungen

Item	h^{2a}		$1-h^{2b}$	λ^c	
	Anfänglich	Extraktion		Faktor 1: <i>Vertrauensbereitschaft</i>	Faktor 2: <i>Misanthropie</i>
v2.11	,161	,373	,627	<,10	,668
v2.13	,168	,276	,724	<,10	,493
v2.15	,241	,340	,660	,610	<,10
v2.32	,264	,377	,623	,619	<,10
v2.37	,223	,291	,709	,259	,339
v2.41	,222	,315	,685	,583	<,10
v2.42	,307	,442	,558	,604	<,10

^a Kommunalität bzw. Anteil der durch den Faktor erklärten Itemvarianz (Bühner (2011), S. 310)

^b *Uniqueness* bzw. Einzigartigkeit; Varianzanteil, den das Item mit keinem der anderen Items teilt (Bühner (2011), S. 300)

^c Faktorladung

(Quelle: eigene Berechnungen)

Die verschiedenen Items tragen unterschiedlich stark zur Varianzaufklärung bei ($h^2 = .276-.442$). Dabei liegt keine der Kommunalitäten unterhalb der $h^2 < .10$ Grenze. Die Varianzaufklärung durch die extrahierten Faktoren kann somit als ausreichend gut bezeichnet werden und erreicht verglichen mit der einfaktoriellen Lösung ein höheres Niveau. Die *Uniqueness*⁶¹ der Items nimmt Werte zwischen .558-.724 an und hat sich im Vergleich mit der einfaktoriellen Lösung verringert. Der Anteil nicht erklärter Varianz (Spezifität und Messfehler) ist allerdings nach wie vor hoch, wobei die mehrdimensionale Faktorenlösung verglichen mit der eindimensionalen Faktorenlösung durchaus ein verbessertes Erklärungsmodell für die Beziehungen unter den einzelnen Items bereitstellt. (Bühner (2011), S. 358)

⁶¹ Vgl. Bühner (2011), S. 300

6. Pilotstudie

Auch der Anteil nicht redundanter Residuen hat sich verringert und liegt nun nur noch bei 1%. Der Modellfit der multifaktoriellen Lösung ist somit besser als der der einfaktoriellen Lösung. (Backhaus/Erichson/Plinke/Weiber (2006), S. 351f; Bühner (2011), S. 366f)

Als Begründung für die ausgeprägte *Uniqueness* der Items können wie bereits zuvor die geringe Augenscheinvalidität der Items in Zusammenhang mit der Präsentation als Blindversuch herangezogen werden.

Als relevant wurden im Folgenden Faktorladungen mit $\lambda \geq .30$ bzw. die jeweils höchste Ladung eines Items gewertet. (Bühner (2006), S. 208f) Die Extraktion ergab daraufhin zwei Faktoren, auf die jeweils 3 bis 4 Items entsprechend der genannten Vorgaben luden. Dabei ergaben sich keinen nennenswerten Doppelladungen, sodass von einer Einfachstruktur ausgegangen werden kann. (Bühner (2011), S. 336, S. 350; Moosbrugger/Kelava (2012), S. 332)

Die Ladungen der Faktor 1: „*Vertrauensbereitschaft*“ zugeordneten Items nehmen Werte $\lambda = .583-.619$ an, wobei Item v2.32 das zugehörige Marker-Item darstellt. Auf Faktor 2: „*Fairness*“ hochladenden Items erreichen Werte von $\lambda = .339-.668$. Das Marker-Item ist hier v2.11. Verglichen mit der multifaktoriellen Lösung der Objektiv-Skala gibt es einige Verschiebungen hinsichtlich der Item-Faktor-Zuordnung. Zunächst sind 7 Items nicht mehr in der Kurzskaala enthalten. Daneben entspricht Faktor 2 der Kurzskaala im Wesentlichen Faktor 3 der Gesamtskala, lediglich Item v2.17 wurde bereits im Zuge der vorgeschalteten Itemanalyse eliminiert und fand deshalb keinen Eingang in die hier durchgeführte Hauptachsenanalyse. Faktor 1 der Kurzskaala setzt sich aus Items des ersten und zweiten Faktors der Gesamtskala zusammen. Auch dieser Umstand ist auf die Eliminierung von Items im Zuge der Itemanalyse zurückzuführen.

Tab. 6.47: Item-Faktor-Zuordnung (Objektiv-Kurzskaala)

Faktor 1: <i>Vertrauensbereitschaft</i>	Faktor 2: <i>Fairness</i>
2.32) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.	2.11) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.
2.15) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.	2.13) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.
2.41) Es ist sicherer, Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.	2.37) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.
2.42) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.	---

Die beiden extrahierten Faktoren weisen einen starken positiven Zusammenhang auf ($r = .621$), was auf einen gemeinsamen übergeordneten Faktor hindeutet. Eine nachfolgende Hauptachsenanalyse höherer Ordnung zur Prüfung dieser Annahme erweist sich allerdings nicht als opportun, da zu wenige Variablen in die Analyse eingehen würden und die herangezogenen Gütekriterien⁶² davon abraten.

Vor dem Hintergrund, dass nach GUADAGNOLI/VELICER eine Faktorstruktur als generalisierend angenommen werden kann, wenn auf mindestens einem der Faktoren vier Variablen Ladungen von $\lambda \geq .60$ aufweisen, ist für die extrahierte Faktorstruktur durchaus eine

⁶² KMO = .50 bzw. MSA-Koeffizient $\geq .50$: Itemauswahl ist nur „schlecht“ („*miserable*“) für eine Faktorenanalyse geeignet. (Backhaus/Erichson/Plinke/Weiber (2006), S. 336; Bühner (2011), S. 347)

entsprechende Tendenz zu erkennen. So weisen drei der vier auf Faktor 1 ladenden Items Ladungen von $\lambda > .60$ auf, während das verbleibende Item nur knapp unterhalb dieser Schwelle rangiert. (Guadagnoli/Velicer (1988), S. 271ff; Klopp (2013), S. 8)

6.2.7 Zusammenfassung

Aus der Item- und Reliabilitätsanalyse resultieren zwei Vertrauensskalen unterschiedlicher Länge, die Objektiv-Skala und die Objektiv-Kurzskala. Während die Objektiv-Skala 14 Items umfasst, sind in der Objektiv-Kurzskala nur 7 Items enthalten. Folgt man der Faktorenlösung erster Ordnung, so ergeben sich für die Objektiv-Skala vier Komponenten und für die Objektiv-Kurzskala zwei Komponenten. Hinsichtlich dieser extrahierten inhaltlichen Dimensionen weisen beide Skalen erhebliche Ähnlichkeit mit der Faktorstruktur der Standardskala (*Trust in People-Scale*) auf. Beide Skalen konvergieren in zweiter Instanz in einem einzelnen Faktor, der als „Generalisiertes Vertrauen“ identifiziert wurde. Es kann somit davon ausgegangen werden, dass sowohl die Objektiv-Skala als auch die Objektiv-Kurzskala vergleichsweise homogene Skalen darstellen, mit deren Hilfe vornehmlich ein einzelnes theoretisches Konstrukt erfasst wird.

7. Hauptstudie

7.1 Methodik

7.1.1 Forschungsdesign

7.1.1.1 Erhebungskonzept

Die Hauptstudie beinhaltet ein Surveyexperiment, dessen Erhebungskonzept eine Kombination aus *within-subject*- und *between-subject*-Design darstellt und somit in Teilen einem *split-ballot*-Design⁶³ entspricht. Im Allgemeinen findet innerhalb von Surveyexperimenten eine systematische Variation einzelner oder mehrerer Elemente eines Fragebogens statt, sodass die Probanden im Rahmen der experimentellen Anordnung zwei oder mehr Varianten eines Fragebogens bearbeiten. Solche Variationen können dabei beispielsweise die Fragereihenfolge oder die Formulierung der Fragen betreffen. Im Rahmen der Analyse werden dann die verschiedenen Varianten des Fragebogens gegenübergestellt und auf Unterschiede im Antwortverhalten der Versuchspersonen hin untersucht. Neben inhaltlichen Aspekten werden mit Hilfe von Surveyexperimenten auch methodische Fragestellungen beforscht. Oft findet auch eine Vermischung beider Facetten statt. (Faas (2009), S. 82; Gaines/Kuklinski/Quirk (2007), S. 3ff; Nock/Guterbock (2010), S. 838ff; Sniderman/Grob (1996), S. 379f) TOURANGEAU unterscheidet im Falle von methodischen Fragen zwei Forschungsziele: (1) *praxis-zentriert*: Auswahl der am besten geeigneten bzw. am besten funktionierenden Fragebogenvariante oder die Bestimmung der Auswirkungen des Einsatzes eines neuen Fragebogens in Kontrast zu einem bestehenden Fragebogen (2) *theoretisch-orientiert*: Identifikation von allgemeinen Charakteristika oder Prinzipien, die zu Items mit einer höheren Reliabilität führen oder ehrlichere Antworten auf Items bewirken. (Tourangeau (2004), S. 222f)

Auch für das vorliegende Surveyexperiment standen methodische Aspekte im Vordergrund. So war es das Ziel zu prüfen, inwiefern sich Messunterschiede zwischen einer Fragebatterie, die nach den Vorgaben für einen Hybriden aus direkten und indirekten Messverfahren erstellt wurde, und einer klassischen Fragebatterie ergeben. Dabei erwiesen sich insbesondere *split-ballot*-Experimente als gut geeignet, „to indicate whether the variation in the method causes differences in the response distribution“ (Saris/Satorra/Coenders (2004), S. 311). Zieldimension beider Skalen war das Konstrukt „Generalisiertes Vertrauen“. Dabei stimmte die Objektiv-Variante mit der Vertrauensskala überein, die im Rahmen der Pilotstudie entwickelt wurde, während die Standard-Variante der Standardskala zur Erfassung Generalisierten Vertrauens (*Trust in People-Scale*) entsprach.

Im Rahmen eines *within-subject*-Designs werden die Probanden mehr als einer bzw. allen möglichen Versuchsbedingungen ausgesetzt. Vorteil einer solchen experimentellen Anordnung ist es, dass festgestellt werden kann, wie sich das Verhalten der Probanden je nach Treatment verändert. So werden jeder Ausprägung der unabhängigen Variablen alle Versuchspersonen zugeordnet und auf diese Weise personenbezogene Störvariablen konstant gehalten bzw. parallelisiert und damit ausgeschlossen. (Charness/Gneezy/Kuhn (2012), S. 1f; Erlebacher (1977), S. 212; Heller (2012), S. 183, S. 185; Huber (2009), S. 172) Gleichzeitig ergibt sich durch die wiederholte Messung an den gleichen Personen die Gefahr der Konfundierung der

⁶³ „Split-ballot experiments randomly divide respondents into one or more treatment groups and a control group.“ (Nock/Guterbock (2010), S. 840) Diesem Aufbau entsprechen insbesondere die als *between-subject*-Design angeordneten Versuchsgruppen.

7. Hauptstudie

verschiedenen Versuchsbedingungen. Typische Validitätsbedrohungen bei experimentellen Designs mit Messwiederholung sind hierbei sogenannte *Positions-* bzw. *Reihenfolge*-Effekte und sogenannte *Carry-over*-Effekte. Im Allgemeinen spielen bei *within-subject*-Designs vornehmlich Effekte der experimentellen Situation eine Rolle, da personengebundene Störvariablen aufgrund der Messwiederholung konstant gehalten werden. (Greenwald (1976), S. 318; Heller (2012), S. 185; Huber (2009), S. 171ff; Shadish/Cook/Campbell (2002), S. 109)

Positionseffekte stellen Störvariablen dar, die sich aus der Reihenfolge der Präsentation der experimentellen Bedingungen ergeben. Beispiele für solche Effekte sind Lern- und Ermüdungseffekte sowie Sensibilisierung, wobei in der Regel die zuletzt vorgegebenen Versuchsbedingungen stärker durch Verzerrungen dieser Art betroffen sind. Eine Kontrolle von Positionseffekten kann durch die Realisierung verschiedener Bearbeitungsreihenfolgen bewerkstelligt werden. Hierzu existieren prinzipiell zwei Strategien: (a) Bei der *Randomisierung* durchläuft jeder der Versuchspersonen eine zufällige Abfolge der Versuchsbedingungen. Die Methode des (b) *Ausbalancierens* differenziert sich in (b1) *vollständiges Ausbalancieren* und (b2) *unvollständiges Ausbalancieren* aus. Im Zuge des *vollständigen Ausbalancierens* werden alle möglichen Reihenfolgen der Versuchsbedingungen realisiert und jeder der Reihenfolgen ein bestimmter Anteil von Probanden zugeordnet. Das *unvollständige Ausbalancieren* umfasst eine Reihe verschiedener Techniken (Spiegelbildmethode, Methode des Lateinischen Quadrats), die im Wesentlichen darauf reduziert werden können, dass nur eine Teilmenge aller möglichen Treatment-Abfolgen im Rahmen des Experimentes umgesetzt werden bzw. nicht jede Versuchsperson alle möglichen Abfolgen bearbeitet. (Abdi/Edelman/Valentin/Dowling (2009), S. 345; Bortz/Döring (2016), S. 550; Heller (2012), S. 185; Huber (2009), S. 171ff)

Ein *Carry-over*-Effekt bezeichnet die inhaltliche Beeinflussung einer experimentellen Bedingung durch ein vorangegangenes Treatment bzw. die Persistenz des Effektes eines Treatments zum Zeitpunkt der Messung des Effektes eines nachfolgenden Treatments. Mit anderen Worten: Zwischen zwei Versuchsbedingungen besteht aufgrund der Reihenfolge bzw. Abfolge der Versuchsbedingungen eine inhaltliche Abhängigkeit. Dabei spielt die absolute Position eines Treatments in der Abfolge der Treatments eine untergeordnete Rolle. Viel wichtiger ist, welche spezielle Versuchsbedingung dem Treatment vorausgegangen ist. Zur Kontrolle bzw. Minimierung von *Carry-over*-Effekten werden eine Reihe von Maßnahmen vorgeschlagen, die von einem ausreichend großen zeitlichen Abstand zwischen den Treatments über eine Umgestaltung der experimentellen Bedingungen zur Eliminierung der Ursache des Effektes bis hin zur Umsetzung des Experimentes in Form eines *between-subject*-Design reichen. (Abdi/Edelman/Valentin/Dowling (2009), S. 345; Greenwald (1976), S. 318; Heller (2012), S. 185f; Huber (2009), S. 173; Kirk (2013), S. 703) Kausalschlüsse auf der Basis einer solchen Versuchsanordnung sind somit nur dann zulässig, wenn die Bearbeitung der verschiedenen experimentellen Instanzen unabhängig voneinander geschieht bzw. eine ein- oder wechselseitige Beeinflussung des Antwortverhaltens der Probanden durch die verschiedenen Versuchsbedingungen ausgeschlossen werden kann. (Charness/Gneezy/Kuhn (2012), S. 1)

Weitere Quellen systematischer Verzerrungen bei experimentellen Designs mit Messwiederholung stellen zwischenzeitliches Geschehen („*history*“), Reifungsprozesse („*maturation*“) und systematische Ausfälle („*mortality*“) dar. Zwischenzeitliches Geschehen beschreibt dabei sämtliche Ereignisse, die zwischen Pre- und Posttest zusätzlich zum Treatment Einfluss auf die abhängige Variable nehmen können. Reifungseffekte beschreiben intrapersonelle und natürliche Veränderungen der Versuchspersonen während der experimentellen Untersuchung, die losgelöst vom Einfluss des Treatments stattfinden.

7. Hauptstudie

Ausfalleffekte ergeben sich, wenn Probanden nur am Pretest, nicht aber am Posttest teilnehmen. Da es sich bei den Ergebnissen von Pre- und Posttest um Messwertpaare handelt, können die Resultate dieser Probanden nicht verwendet werden. Die Folge ist eine Verringerung der in die Untersuchung eingehenden Fallzahl. (Reiß/Sarris (2012), S. 89; Schnell/Hill/Esser (2008), S. 217ff; Shadish/Cook/Campbell (2002), S. 56ff)

Einem *within-subject*-Design entsprechen in der Hauptstudie jeweils die *Gruppe 1: Objektiv/Standard* sowie die *Gruppe 2: Standard/Objektiv*. Beide Versuchsgruppen entsprechen insofern einem experimentellen Design mit Messwiederholung, als dass die Probanden in jeder der Gruppen sowohl die Objektiv-Skala als auch die Standard-Skala bearbeiteten. Das bedeutet: Die individuelle Ausstattung der Probanden mit Generalisiertem Vertrauen wird zweimal zu unterschiedlichen Messzeitpunkten anhand unterschiedlicher Vertrauensskalen erfasst. Die Auswahl dieser Form von experimentellem Design erfolgte, um Unterschiede im Antwortverhalten auf der Objektiv-Skala im Kontrast zur der Standard-Skala aufzudecken, wobei der Einfluss personengebundener Störvariablen konstant gehalten werden sollte. Die Kontrolle möglicher Positionseffekte erfolgte dabei anhand der Methode des Lateinischen Quadrats. So wurden in den beiden Gruppen wechselseitig zueinander invers verlaufende Abfolgen der Messbedingungen zur Erfassung Generalisierten Vertrauens umgesetzt. Dabei bearbeitete *Versuchsgruppe 1* zunächst die Objektiv-Skala und im Anschluss die Standard-Skala, während für *Versuchsgruppe 2* die umgekehrte Bearbeitungsreihenfolge gilt. *History*-, *Maturation*- und *Mortality*-Effekten wurde dadurch begegnet, dass die Zeitspanne zwischen erster und zweiter Befragungsrunde möglichst minimal gestaltet wurde bzw. der Posttest jeweils im direkten Anschluss an den Pretest erfolgte. Zur Kontrolle von Carry-over-Effekten wurde das Surveyexperiment zusätzlich in Form eines *between-subject*-Designs durchgeführt.

Ein *between-subject*-Design setzt jeden Probanden nur einem Treatment aus. Die Anzahl der zu realisierenden Versuchsgruppen entspricht somit der Anzahl der zu realisierenden Versuchsbedingungen. Kausalschlüsse lassen sich bei dieser Art von experimenteller Anordnung durch den Vergleich des Verhaltens der Probanden unter der einen Versuchsbedingung mit dem Verhalten der Probanden, die einer anderen experimentellen Kondition ausgesetzt waren, ziehen. (Charness/Gneezy/Kuhn (2012), S. 1; Heller (2012), S. 183) Validitätsbedrohungen ergeben sich bei Versuchsplänen ohne Messwiederholung vornehmlich aus personengebundenen Störvariablen. Diese resultieren daraus, dass sich die Zusammensetzung der verschiedenen Experimentalgruppen nicht nur bezüglich der unabhängigen Variablen unterscheidet, sondern auch hinsichtlich anderer für die Fragestellung relevanter Merkmale. (Heller (2012), S. 183) Kontrolltechniken für Verzerrungen solcher Art stellen die Randomisierung, das Konstanthalten und die Parallelisierung bzw. das Matching dar. (Heller (2012), S. 183f; Huber (2009), S. 105ff)

Strenggenommen lässt sich im Rahmen des hier vollzogenen Survey-Experimentes jede Versuchsgruppe in Relation zu einer anderen Versuchsgruppe als *between-subject*-Design anordnen. So stellen die Erhebungen in den unterschiedlichen Versuchsgruppen wechselseitig voneinander unabhängige Messungen dar. Grundvoraussetzung bei der Gegenüberstellung der einzelnen Versuchsgruppen ist allerdings, dass immer Objektiv-Skala und Standard-Skala kontrastiert werden. Dies ist insbesondere für den Vergleich mit den Messungen Generalisierten Vertrauens zu beachten, die aus den *Versuchsgruppen 1* und *2* resultieren, in denen beide Skalentypen zum Einsatz kamen. Genuin als *between-subject*-Design konzipiert wurde ausschließlich die Kontrastierung von *Versuchsgruppe 3: Objektiv* und *Versuchsgruppe 4: Standard*.

7. Hauptstudie

Einflüsse der oben angesprochenen personengebundenen Störvariablen wurden im Rahmen des *between-subject*-Design durch eine *Matching* der Probanden bzw. eine Parallelisierung der Versuchsgruppen anhand des Geschlechterverhältnisses und der Altersstruktur versucht zu neutralisieren.

Generell sind im Rahmen experimenteller Anordnungen auch sogenannte Versuchsleiter-Erwartungseffekte („*Rosenthal*-Effekt“, „*Pygmalion*-Effekt“) und reaktive Effekte der experimentellen Situation („*Hawthorne*-Effekt“) zu beachten. Ein *Rosenthal*- bzw. *Pygmalion*-Effekt ergibt sich daraus, dass sich die Probanden in ihrem Verhalten an den Ergebniserwartungen des Versuchsleiters orientieren. Dies kann u.a. geschehen, wenn der Versuchsleiter durch verbale Äußerungen, Mimik oder Körpersprache Hinweise auf die zu prüfende Hypothese macht oder wenn den Probanden die fragliche Hypothese bekannt ist. (Bortz/Döring (2016), S. 101; Diekmann (2011), S. 66, S 624f; Eid/Gollwitzer/Schmitt (2015), S. 54; Huber (2009), S. 184ff) Um diese Form selbsterfüllender Prophezeiung zu vermeiden, wurden die Probanden im Vorhinein der Studie nicht über die tatsächlichen Forschungsziele aufgeklärt. Außerdem wurden die Probanden im Debriefing dazu angehalten, Stillschweigen über die Zielsetzung der Studie zu bewahren, bis die Erhebung vollständig abgeschlossen sein würde.

Als *Hawthorne*-Effekt werden Verhaltensänderungen der Probanden bezeichnet, die auf das Bewusstsein, Gegenstand einer wissenschaftlichen Untersuchung bzw. einer Testsituation zu sein, zurückzuführen sind und dabei unabhängig von der Wirksamkeit eines Treatments auftreten. (Bortz/Döring (2016), S. 101; Diekmann (2011), S. 341; Eid/Gollwitzer/Schmitt (2013), S. 53) Veränderungen im Antwortverhalten der Probanden aufgrund eines *Hawthorne*-Effektes können möglicherweise in den Versuchsgruppen, die einem experimentellen Design mit Messwiederholung folgen (*within-subject*-Design), eine Rolle spielen. Überlegungen seitens der Probanden, welche inhaltliche Verbindung zwischen Pretest und Posttest bestehen, könnten Verzerrungen in den Messergebnissen des jeweiligen Posttests bewirken. Eine Kontrolle dieses Effektes kann dabei erst ex-post erfolgen und ist in die Kontrollstrategie für *Positions*-, *Reihenfolge*- und *Carry-over*-Effekte eingebettet bzw. entspricht dieser.

Daneben können auch die sogenannten *demand characteristics* bzw. die Anforderungsmerkmale der experimentellen Situation („*Erwartungseffekt*“) Auswirkungen auf das Antwortverhalten der Probanden haben. So versuchen die Probanden in der Regel anhand der experimentellen Prozeduren Rückschlüsse auf den Zweck des Experimentes bzw. die zugrundeliegende Hypothese zu ziehen, um ihr Antwortverhalten entsprechend anzupassen. Bei der Hypothesenbildung durch die Probanden handelt es sich allerdings um einen individuell höchst subjektiven Prozess, an dessen Ende nicht unbedingt die bezüglich des Experimentes tatsächlich richtige Hypothese erraten wird. Entsprechend können auch „falsche“ Hypothesen handlungsleitend für die Versuchspersonen werden, sodass bereits eine fehlerhafte Interpretation des Untersuchungsziels zu Verzerrungen im Untersuchungsergebnis führen kann. Im Anschluss an diese Interpretationsleistung ergeben sich prinzipiell vier mögliche Handlungsmuster für das Verhalten der Versuchspersonen: (1) „*good subject*“: die Versuchspersonen zeigen hypothesenkonformes Verhalten, damit sich die Annahmen des Experimentes bestätigen; (2) „*negativistic subject*“: die Versuchspersonen zeigen ein Antwortverhalten, dass der Hypothese des Experimentes entgegengesetzt ist, um diese zu widerlegen; (3) „*evaluation apprehension*“: die Versuchspersonen möchten eine möglichst positive Bewertung seitens des Versuchsleiters erreichen; (4) „*faithful subject*“: die Versuchspersonen lassen sich nicht durch die wahrgenommenen *demand characteristics*

7. Hauptstudie

beeinflussen und handeln auf Basis der für das Experiment vorgegebenen Instruktionen. (Bungard (1984), S. 116ff; Charness/Gneezy/Kuhn (2012), S. 2; Huber (2009), S. 192; Kirk (2013), S. 20; Orne (1962), S. 778ff; Orne (1969), S. 147ff; Rosenthal/Rosnow (2008), S. 216ff; Schnell/Hill/Esner (2011), S. 220; Shadish/Cook/Campbell (2002), S. 77f)

Die Möglichkeiten zur Kontrolle von *Demand*-Effekten sind dadurch beschränkt, dass kein Experiment frei von derart gelagerten Einflüssen sein kann, da auch Hinweise, die nichts mit der Fragestellung einer Studie zu tun haben für die Konstruktion einer Hypothese herangezogen werden. (Bungard (1984), S. 117):

„It should be clear that demand characteristics cannot be eliminated from experiments; all experiments will have demand characteristics, and these will always have some effect.“ (Orne (1962), S. 779)

Relevant werden solche Effekte allerdings erst dann, wenn sie mit dem Effekt der unabhängigen Variablen konfundiert sind. Da die Funktion Objektiver Tests im Wesentlichen von der erfolgreichen Verschleierung der Messintention abhängig ist, entsteht diese Problematik im Rahmen der Hauptstudie, wenn die für die Objektiv-Items angestrebte geringe Augenscheinvalidität bereits durch die Testinstruktionen aufgedeckt würde. Eine Aufklärung der Probanden über die Zieldimension der Objektiv-Skala hätte ein Priming der Probanden auf die Zieldimension „Generalisiertes Vertrauen“ zur Folge, was eine Verzerrung der Ergebnisse der Objektiv-Messung nach sich ziehen würde. Spontane unverzerrte Reaktionen auf die Itemformulierungen würden unmöglich, was dem oben beschriebenen Erwartungseffekt entspricht. Um eine unvoreingenommene Bearbeitung der Objektiv-Skala durch die Probanden zu gewährleisten, wurde diesen zur Ablenkung von der wahren Zielsetzung der Studie eine Coverstory präsentiert. Hiernach misst die Objektiv-Skala nicht „Generalisiertes Vertrauen“, sondern erfasst die Perzeption anderer Menschen durch die jeweiligen Probanden. Im Anschluss an die Erhebung fand ein Debriefing der Probanden statt, das über den wahren Zweck der Befragung aufklärte. (Bortz/Döring (2016), S. 197; Bröder (2011), S. 149; Shadish/Cook/Campbell (2002), S. 78)

Alternativ kann die Anordnung der verschiedenen Versuchsgruppen des Surveyexperimentes auch als eine Adaption des Solomon-Vier-Gruppen-Designs betrachtet werden. Das Solomon-Vier-Gruppen-Design stellt eine Vermischung des Two-Group-Pretest-Posttest-Designs mit dem Posttest-Only-Design dar und wird insbesondere dazu eingesetzt, sogenannte Pretesteffekte aufzudecken. Ein Pretesteffekt beschreibt eine Beeinflussung der abhängigen Variable durch eine Vorher-Messung und resultiert in einer Verfälschung der Posttest-Ergebnisse. Es findet somit eine Konfundierung von Pretest und Treatment statt, in deren Folge man nicht mehr unterscheiden kann, ob die Veränderungen innerhalb der abhängigen Variablen auf den Einfluss des Pretests oder den Einfluss des Treatments zurückzuführen ist bzw. beides gemeinsam die Variation in der abhängigen Variable bewirkt hat. (Bortz/Döring (2016), S. 731; Braver/Braver (1988), S. 150; Diekmann (2011), S. 344; Schnell/Hill/Esner (2008), S. 224f; Shadish/Cook/Campbell (2002), S. 60)

Tab. 7.1: Solomon-Vier-Gruppen-Design

Design	Gruppe		Pretest	Treatment	Posttest
Two-Group-Pretest-Posttest-Design	Experimentalgruppe 1	R	O ₁	X	O ₂
	Kontrollgruppe 1	R	O ₃		O ₄
Posttest-Only-Design	Experimentalgruppe 2	R		X	O ₅
	Kontrollgruppe 2	R			O ₆

7. Hauptstudie

Der wesentliche Unterschied zwischen Solomon-Vier-Gruppen-Design und dem im Rahmen der Hauptstudie durchgeführten Survey-Experiment besteht darin, dass in Letzterem keine klassischen Kontrollgruppen existieren. Dies ergibt sich aus dem Ziel der Untersuchung, welches darin besteht, den Effekt zweier verschiedener Messmethoden auf die Ausprägung einer unabhängigen Variablen zu kontrastieren. Als unabhängige Variable des Surveyexperimentes kann somit die Messmethode verstanden werden, während die abhängige Variable das individuelle Ausmaß Generalisierten Vertrauens darstellt. Das Treatment des Experimentes bestand demnach aus der Variation der Messmethode innerhalb der verschiedenen Gruppen. Ausgehend hiervon stellte jede der im Rahmen des Surveyexperimentes realisierten Versuchsbedingungen gleichermaßen eine Versuchsgruppe wie eine Kontrollgruppe für die übrigen Versuchsbedingungen dar. Dies ergibt sich daraus, dass in jeder der Versuchsgruppen mindestens eine der Messmethoden zur Erfassung Generalisierten Vertrauens zum Einsatz kam. Eine klassische Kontrollmessung im Sinne der Erfassung der Ausprägung der abhängigen Variablen ohne die Beeinflussung der Probanden durch ein Treatment ist aufgrund der Anlage des Surveyexperimentes unmöglich, da die Art und Weise der Messung das Treatment darstellt. Eine klassische Kontrollgruppe, d.h. Probanden, die keinem Treatment ausgesetzt wurden, würde hier einer Gruppe entsprechen, innerhalb der keinerlei Messung durchgeführt würde. Dies hätte allerdings auch zur Folge, dass innerhalb dieser Gruppe keine Information hinsichtlich der Ausprägung der abhängigen Variable zugänglich wäre bzw. erfasst würde. Ein solche Gruppe kann demnach als überflüssig angesehen werden.

Tab. 7.2: Versuchsplan Hauptstudie

Versuchsgruppe	Zuordnung	Pretest	„Treatment“	Posttest	Experimentelles Design
Objektiv/Standard	R ^a	Fragebogen (Objektiv-Variante)	Standard vs. Objektiv Veränderung des Itemdesigns	Fragebogen (Standard-Variante)	Within-subject
Standard/Objektiv	R ^a	Fragebogen (Standard-Variante)	Standard vs. Objektiv Veränderung des Itemdesigns	Fragebogen (Objektiv-Variante)	Within-subject
Objektiv	R ^a		Standard od. Objektiv	Fragebogen (Objektiv-Variante)	Between-subject
Standard	R ^a		Standard od. Objektiv	Fragebogen (Standard-Variante)	

^a „Random“ (per Zufall)

Für die konkrete Umsetzung der Erhebung ergab sich folgendes Muster. Die im *within-subject*-Design angeordneten *Versuchsgruppen 1* und *2* beinhalteten jeweils zwei voneinander zeitlich getrennte Messzeitpunkte. Zu beiden Messzeitpunkten wurde den Probanden jeweils eine der beiden Messvarianten zur Bearbeitung vorgelegt. Nachdem dieser Fragebogen ausgefüllt und an den Versuchsleiter zurückgegeben worden war, wurde der jeweilige Pendant-Fragebogen zur Bearbeitung freigegeben. So bekamen die Probanden im Rahmen von *Versuchsgruppe 1* zunächst die Objektiv-Variante und erst in der zweiten Befragungsrunde die Standard-Variante präsentiert. Für *Versuchsgruppe 2: Standard/Objektiv* ergab sich die umgekehrte Reihenfolge.

Die im *between-subject*-Design angeordneten *Versuchsgruppen 3* und *4* beinhalteten jeweils nur einen Messzeitpunkt. So bekamen die Probanden in *Versuchsgruppe 3: Objektiv* ausschließlich die Objektiv-Variante präsentiert, während die Probanden in *Versuchsgruppe 4: Standard* ausschließlich die Standard-Variante bearbeiteten.

7.1.1.2 Stichprobe

Wie bereits in der Pilotstudie handelt es sich auch bei den für die Hauptstudie realisierten Stichproben um sogenannte *convenience sample*, die sich aus Studierenden der Politikwissenschaft und Soziologie der Universitäten Regensburg und Innsbruck zusammensetzen.

Die Verwendung willkürlicher Stichproben im Rahmen eines Surveyexperimentes ist insofern gerechtfertigt, als dass die Zielsetzung eines Experimentes die Prüfung von Hypothesen darstellt. Im Gegensatz zu einer repräsentativen Befragung steht dabei nicht die Generalisierung der Ergebnisse in Vordergrund. (Leiner (2014), S. 5) Darüber hinaus stellt die Verwendung studentischer Samples im Rahmen experimenteller Untersuchungen eine gängige Praxis dar. (Eid/Gollwitzer/Schmitt (2015), S. 92; Reiß/Sarris (2012), S. 114) Daneben spielten auch forschungspragmatische Gründe eine Rolle für die Verwendung von *convenience samples*. So ist die Ziehung von Zufallsstichproben in der Regel mit erheblichen finanziellen Aufwendungen verbunden, die im Rahmen der vorliegenden Studie nicht geleistet werden konnten. Sogenannte Ad-hoc-Stichproben stellen hierbei eine kostengünstige Alternative dar. (Bortz/Döring (2016), S. 305ff; Huber (2009), S. 116)

Tab. 7.3: Zusammensetzung der Versuchsgruppen

Versuchsgruppe	N	Geschlecht		Alter (Jahre)			
		♀	♂	μ	σ	Min	Max
Objektiv/Standard	50	28	22	23,32	2,43646	19	30
Standard/Objektiv	50	26	24	22,26	2,66351	18	32
Objektiv	52	32	20	22,8269	3,24020	19	35
Standard	52	28	24	22,3269	2,46323	19	29
Gesamt	204	114	90	22,6814	2,73662	18	35

Da vier Versuchsgruppen realisiert werden sollten, wurden insgesamt vier voneinander unabhängige Stichproben mit N=50 bzw. N=52 gezogen. In Summe gehen 204 Fälle in das Surveyexperiment ein. Auf die gesamte Studie bezogen sind Frauen leicht überrepräsentiert. Mit Ausnahme von *Gruppe 3: Objektiv* ergeben sich für die einzelnen Versuchsgruppen allerdings verhältnismäßig ausgeglichene Geschlechterverhältnisse. Über alle Versuchsgruppen hinweg lässt sich eine homogene Altersstruktur feststellen. Dies war aufgrund der Verwendung studentischer Samples zu erwarten.

7.1.1.3 Feldzugang

Bei der Organisation des Feldzuganges wurden verschiedene Strategien verfolgt. So erforderten insbesondere die nach dem *within-subject*-Design konzipierten Versuchsgruppen ein besonderes Maß an Kontrolle, mussten bei diesen Gruppen doch beide Versuchsbedingungen realisiert werden, wobei eine wechselseitige Beeinflussung der Messvarianten verhindert werden sollte. Für die Erhebungen im Rahmen des *within-subject*-Designs galt dabei unabhängig von der Rekrutierungsstrategie, dass die Probanden zunächst nur eine der Messvarianten präsentiert bekamen. Nach Bearbeitung des jeweiligen Fragebogens wurde dieser wieder eingesammelt und eine weitere Befragungsrunde mit der zweiten Messvariante durchgeführt. Für *Versuchsgruppe 1: Objektiv/Standard* bedeutete dies, dass die Probanden zunächst die Objektiv-Variante bearbeiteten und danach die Standard-Variante. Für *Versuchsgruppe 2: Standard/Objektiv* galt die umgekehrte Reihenfolge.

Es wurde zunächst versucht, Studierende per Rundmail zu rekrutieren. Hierzu standen die Mail-Adressen sämtlicher Studierender der Politikwissenschaft der Universität Regensburg zur Verfügung. Im Rahmen der Rundmail wurden die Studierenden zu einem Surveyexperiment mit dem Titel „Perzeption/Wahrnehmung anderer Menschen“ eingeladen, wobei die Studierenden aus drei verschiedenen Erhebungszeitpunkten wählen konnten und sich zum jeweils ausgewählten Termin in einem speziellen Raum einfinden sollten. Auf diese Weise sollte ein Höchstmaß an Kontrolle über die experimentelle Situation gewährleistet werden. Im Rahmen dieser Maßnahme konnten allerdings ausschließlich ausreichend Probanden rekrutiert werden, um *Versuchsgruppe 1* zu besetzen. Im weiteren Verlauf des Surveyexperimentes wurde daher die Strategie zur Rekrutierung von Versuchspersonen verändert.

Um eine ausreichend große Fallzahl zu erreichen, wurde alternativ festgelegt, Probanden direkt in Lehrveranstaltungen der Fächer Politikwissenschaft und Soziologie an den Universitäten Regensburg und Innsbruck anzuwerben. Um eine Konfundierung der verschiedenen Messvarianten zu vermeiden, wurde in den Lehrveranstaltungen dabei jeweils nur eine Versuchsbedingung realisiert. Das bedeutet: In den einzelnen Lehrveranstaltungen wurden immer nur Probanden für eine spezielle Versuchsgruppe rekrutiert. Außerdem durften Studierende, die bereits an einer Fragerunde teilgenommen hatten, kein weiteres Mal am Surveyexperiment teilnehmen. Dies wurde durch eine entsprechende Eingangsfrage im Vorhinein der jeweiligen Befragung sichergestellt.

Der Zugang zu den Lehrveranstaltungen wurde über die jeweils zugeordneten Lehrpersonen hergestellt. Um die Teilnahmebereitschaft der Studierenden zu erhöhen bzw. das Problem des *Unit-Nonresponse* zu verringern, wurden die Befragungen außerdem zu Beginn der einzelnen Lehrveranstaltungen durchgeführt.

7.1.1.4 Operationalisierung

Im Rahmen des Survey-Experimentes der Hauptstudie kommt im Wesentlichen ein einzelner Fragebogen zum Einsatz, der sich allerdings aus zwei Teilen zusammensetzt, die je nach Versuchsgruppe in anderer Art und Weise kombiniert wurden. Dabei waren im Fragebogen für die *Versuchsgruppen 1* und *2* (*Objektiv/Standard* bzw. *Standard/Objektiv*) insgesamt 41 Items enthalten, während der Fragebogen für *Versuchsgruppe 3* (*Objektiv*) 37 Items enthielt und der Fragebogen für *Versuchsgruppe 4* (*Standard*) aus 24 Items bestand. Die Reduktion der Itemanzahl in den *Versuchsgruppen 3* und *4* ergab sich daraus, dass in diesen Gruppen jeweils nur eine der beiden Vertrauensskalen präsentiert wurde. Alle in der Hauptstudie verwendeten Items entsprachen einem geschlossenen Frageformat.

Der Fragebogen erfasst neben Generalisiertem Vertrauen außerdem eine Auswahl von Korrelaten, die mit diesem Konstrukt assoziiert sind, sowie die Tendenz der Probanden sozialerwünscht zu antworten.

Bei den in den *Versuchsgruppen 1 bis 3* verwendeten Fragebogen wurde den Probanden ein Messkontext suggeriert, der zur Verschleierung der Messintention der Objektiv-Skala beitragen sollte. So wurde diesen im Eingangstext mitgeteilt, dass die Befragung vornehmlich erfassen möchte, wie die Probanden andere Menschen wahrnehmen bzw. ob sie anderen Menschen eher positive oder negative Eigenschaften zuschreiben. Da in *Versuchsgruppe 4* ausschließlich die Standard-Skala bearbeitet werden sollte, entfiel dort die Maskierung der Messintention. Stattdessen wurden die Probanden darüber aufgeklärt, dass der Zweck der Befragung in der

Erfassung ihres individuellen Vertrauens bzw. Misstrauens gegenüber anderen Menschen im Allgemeinen besteht.

Nachfolgend sollen die einzelnen Items bzw. Itembatterien entsprechend der Chronologie des in *Versuchsgruppe 1* eingesetzten Fragebogens vorgestellt werden.⁶⁴

7.1.1.4.1 Glück und subjektives Wohlbefinden/Zufriedenheit

Die Items zur Abfrage der tagesaktuellen und allgemeinen Lebenszufriedenheit der Probanden entsprechen den bereits im Rahmen der Pilotstudie verwendeten Items und es gelten die gleichen Annahmen hinsichtlich der theoretischen Verflechtung von subjektiver Zufriedenheit und Generalisiertem Vertrauen. So wirkt sich ein ausgeprägtes subjektives Wohlbefinden positiv auf Einstellungen gegenüber Fremden, die Weltwahrnehmung und die Kooperationsbereitschaft aus. Diese Aspekte haben wiederum einen förderlichen Effekt auf Generalisiertes Vertrauen. Personen, die Erfolg im Leben haben, sind vertrauensvoller, da sie von ihren Mitmenschen meist freundlicher behandelt werden und oft über größere finanzielle und materielle Ressourcen verfügen, sodass Vertrauensenttäuschungen für diesen Personenkreis ein geringeres Risiko darstellen. (Delhey/Newton (2002), S. 3; Delhey/Newton (2004), S. 153)

7.1.1.4.2 Objektiv-(Kurz)Skala

Die Objektiv-Skala bzw. Objektiv-Kurzskala entspricht der im Rahmen der Pilotstudie entwickelten Vertrauensskala. Wie sich infolge faktorenanalytischer Untersuchungen zeigte, entspricht ihre Faktorstruktur in weiten Teilen der Faktorstruktur der *Trust in People-Scale*. Hiernach setzt sich die Objektiv-Skala aus insgesamt vier Dimensionen (*Vertrauensbereitschaft*, *Misanthropie*, *Fairness* und *Hilfsbereitschaft*) zusammen, die in zweiter Instanz in einem einzelnen Faktor „Generalisiertes Vertrauen“ konvergieren. Die Objektiv-Kurzskala setzt sich dagegen aus nur zwei Faktoren (Vertrauensbereitschaft und Fairness) zusammen, die allerdings stark positiv miteinander korreliert sind, sodass von einem übergeordneten gemeinsamen Faktor „Generalisiertes Vertrauen“ ausgegangen werden kann. Während die Objektiv-Skala 14 Items umfasst, enthält die Objektiv-Kurzskala nur 7 Items, die allerdings auch Bestandteil der Objektiv-Skala sind.

⁶⁴ Vgl. Anhang IV: Fragebogen, IV.c1 Hauptstudie (*Experimentalgruppe 2: Standard/Objektiv*) bis IV.e Hauptstudie (*Experimentalgruppe 4: Standard*) für die Itemanordnung innerhalb der Versuchsgruppen 2 bis 4

Tab. 7.4: Items der Objektiv-(Kurz)Skala

(1)	2.4) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden. (+)	Items der Objektiv-Kurzskala
(2)	2.5) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen. (+)	
(3)	2.7) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt. (-)	
(4)	2.11) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun. (-)	
(5)	2.12) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden. (+)	
(6)	2.13) Es ist sicherer, Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt. (-)	
(7)	2.14) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern. (-)	
(8)	2.1) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen. (-)	Ergänzung zur Objektiv-Skala
(9)	2.2) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken. (-)	
(10)	2.3) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinandertreffen, ist von andauernden Konflikten geprägt. (-)	
(11)	2.6) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen. (-)	
(12)	2.8) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte. (-)	
(13)	2.9) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden. (-)	
(14)	2.10) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist. (+)	

Hoher Itemscore entspricht starker Ausprägung der Zieldimension, niedriger Itemscore entspricht schwacher Ausprägung der Zieldimension. Es ergeben sich daher die folgenden Codierungen für die Antwortskala:

(+) Itempolung positiv: „Stimme nicht zu“ (1); „Stimme eher zu“ (2); „Stimme teils zu, stimme teils nicht zu“ (3); „Stimme eher zu“ (4); „Stimme zu“ (5)
 (-) Itempolung negativ: „Stimme nicht zu“ (5); „Stimme eher zu“ (4); „Stimme teils zu, stimme teils nicht zu“ (3); „Stimme eher zu“ (2); „Stimme zu“ (1)

(Quelle: Eigene Darstellung)

Um die Messintention der Objektiv-Skala bzw. Objektiv-Kurzskala zu verschleiern, wurde der Itembatterie eine Coverstory vorangestellt. Anstatt darauf hingewiesen zu werden, dass mit den nachfolgenden Items Generalisiertes Vertrauen erfasst werden soll, war nur von Einstellungsstatements die Rede, die allgemein verbreitete Überzeugungen betreffen. Die Probanden sollten angeben, inwieweit sie mit diesen Aussagen übereinstimmen bzw. im Widerspruch stehen.

7.1.1.4.3 Kognitiver Pretest

Auch im Rahmen der Hauptstudie wurde überprüft, ob die Augenscheinvalidität der Items von Objektiv-Skala bzw. Objektiv-Kurzskala bezüglich der Zieldimension „Generalisiertes Vertrauen“ ausreichend niedrig ausfällt, um eine erfolgreiche Verschleierung der Messintention zu gewährleisten. Die hierzu verwendeten Items stimmen in ihrem Inhalt weitestgehend mit den im Kontext der Pilotstudie zum gleichen Zweck verwendeten Items überein, unterscheiden sich allerdings stellenweise im Wortlaut. So sollten die Probanden nun die Nummer des Statements angeben, an der sie die Intention der Objektiv-Skala bzw. Objektiv-Kurzskala erkannt haben. Auf diese Weise konnte die zur Beantwortung dieser Frage benötigte Zeit im Vergleich zur Pilotstudie deutlich verkürzt werden. In der Pilotstudie mussten die Probanden noch das fragliche Item sinngemäß oder wörtlich wiedergeben. Das methodische Vorgehen entspricht wie

bereits in der Pilotstudie einer Adaption der *Think-Aloud-Technik* für standardisierte Fragebogen.

7.1.1.4.4 Radiusproblem

Im Rahmen der Hauptstudie soll u.a. überprüft werden, ob die Objektiv-Skala bzw. Objektiv-Kurzskala einen signifikanten Beitrag zur Lösung des Radiusproblems beisteuert. Zu diesem Zweck wird sowohl im Anschluss an die Objektiv-Skalen als auch im Nachhinein der Standard-Skala gefragt, welchen Personenkreis die Probanden als Referenzgruppe für ihre Urteile auf den genannten Skalen nutzen. Zum Einsatz kommt dabei jeweils das gleiche Item, wie es bereits innerhalb der Pilotstudie verwendet wurde. Die Antwortalternativen dieses Items differenzieren dabei zwischen zwei Vertretern der *ingroup* und drei Vertretern der *outgroup*.

7.1.1.4.5 Soziale Erwünschtheit

Zur Erfassung möglicher Effekte sozial erwünschten Antwortverhaltens und daraus resultierender potentieller Ergebnisverfälschungen werden u.a. sogenannte Kontrollskalen (SD-Skalen) eingesetzt. Während im Rahmen der Pilotstudie noch eine von ZUMA aus der deutschen Version der *Marlowe/Crowne-Scale* (MC-Skala) abgeleitete Kurzskala benutzt wurde, kommt innerhalb der Hauptstudie eine auf das von PAULHUS entwickelte *Balanced Inventory of Desirable Responding* (BIDR) zurückgehende SD-Skala zum Einsatz. (Paulhus (1984); Robinson/Shaver/Wrightsmann (1991), S. 37ff) Diese SD-Skala geht davon aus, dass es sich bei dem Phänomen „Soziale Erwünschtheit“ nicht um ein unidimensionales Konstrukt handelt, sondern zwischen Selbsttäuschung (*self-deceptive enhancement*, SDE) und Fremdtäuschung (*impression management*, IM) unterschieden werden muss. Während es sich bei SDE um eine unbewusste Täuschung handelt, die sich in einer „Tendenz, die Realität in einer optimistischen Weise verzerrt wahrzunehmen“ bzw. „dem Schutz des Selbstbildes und des Selbstwertes“ (Winkler/Kroh/Spiess (2006), S. 3) manifestiert, beschreibt IM eine absichtliche und bewusste Täuschung mit dem Ziel ein möglichst positives Selbstbild nach außen abzugeben. (Musch/Brockhaus/Bröder (2002), S. 121)

Die BIDR-Kurzskala setzt sich aus insgesamt sechs Items zusammen, wobei jeweils drei Items auf die Komponenten Selbsttäuschung und Fremdtäuschung entfallen. Es sind dabei sowohl negativ als auch positiv gepolte Items enthalten. Zur Beantwortung der Aussagen wird eine siebenstufige Ratingskala mit 1 = „trifft überhaupt nicht zu“ bis 7 = „trifft voll zu“ eingesetzt.

Tab. 7.5: BIDR-Kurzskala

Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus.	SDE +
Ich bin mir oft unsicher in meinem Urteil.	SDE -
Ich weiß immer genau, wieso ich etwas mag.	SDE +
Ich habe schon mal zuviel Wechselgeld zurückbekommen und nichts gesagt.	IM -
Ich bin immer ehrlich zu anderen.	IM +
Ich habe gelegentlich mal jemanden ausgenutzt.	IM -

Auswertungsschlüssel:

Als SD-Kennwert wurde der Summenscore über die gesamte Skala verwendet; als SDE- bzw. IM-Kennwert wurde der Summenscore über die jeweilige Teilskala verwendet. Ein hoher Wert bedeutet dabei eine starke Neigung zu sozialerwünschtem Antwortverhalten, während ein niedriger Wert eine geringe Neigung zu sozialerwünschtem Antwortverhalten anzeigt.

(Quelle: Winkler/Kroh/Spiess (2006), S. 15)

Für den Wechsel der SD-Skala von Pilotstudie zu Hauptstudie sind vornehmlich zwei Gründe zu nennen. So handelt es sich bei der *Marlowe/Crowne-Scale* um eine verhältnismäßig alte Skala, deren Itemformulierungen in vielerlei Hinsicht nicht mehr dem Zeitgeist entsprechen. Zudem zeigen Untersuchungen von WINKLER/KROH/SPIESS und MUMMENDEY/EIFLER, dass die gekürzte MC-Skala den Kriterien für Reliabilität und Validität nicht genügt. (Mummendey/Eifler (1993); Winkler/Kroh/Spiess (2006), S. 2, S. 7ff) Der Hauptgrund ist allerdings darin zu finden, dass die Verwendung der BIDR-Kurzskala einen direkten Vergleich mit den Ergebnissen der Studie von NAEF/SCHUPP zu lässt. NAEF/SCHUPP untersuchten u.a., inwiefern soziale Erwünschtheit einen Einfluss auf das Antwortverhalten auf der *Trust in People-Scale* hat, und setzten hierzu die BIDR-Kurzskala als Kontrollskala ein. Dabei fanden sie heraus, dass insbesondere die IM-Komponente eine Rolle für Verzerrungen innerhalb der Vertrauensmessung spielt. (Naef/Schupp (2009), S. 10)

7.1.1.4.6 Korrelate Generalisierten Vertrauens

Variablen, die im Zusammenhang mit Generalisiertem Vertrauen stehen, wurden in erster Linie zur Verifikation der Validität der Objektiv-Skalen in den Fragebogen aufgenommen. Im Speziellen sollte durch die Bestimmung des Vertrauens gegenüber Implementationsinstitutionen sowie der individuellen Eingebundenheit in informelle und formelle soziale Netzwerke die Kriteriumsvalidität des nach den Vorgaben für die Hybridklasse erstellten Untersuchungsinstrumentes bestimmt werden.

Implementationsinstitutionen wie Judikative und Administration sorgen durch die Überwachung und Gewährleistung der Rechtsordnung bzw. institutioneller Regeln für die Aufrechterhaltung sozialer Ordnung. Die daraus resultierende normative Rahmung für das Handeln der Menschen, vereinfacht die Antizipation des Verhaltens der Menschen, was die Vergabe von Vertrauen wahrscheinlicher macht. (Delhey/Newton (2004), S. 155)

Formelle und informelle soziale Netzwerke ermöglichen gleichermaßen Interaktion, Kommunikation und Kooperation von Individuen. Positive Kooperationserfahrungen haben dabei einen förderlichen Effekt auf Generalisiertes Vertrauen. Außerdem begünstigt ein großer Kreis an Freunden und Bekannten sowie eine regelmäßige Interaktion mit diesen Personen die Vertrauensbereitschaft. (Delhey/Newton (2002), S. 6ff; Delhey/Newton (2004), S. 153f)

Um zu vermeiden, dass anhand der Abfrage des Vertrauens gegenüber Implementationsinstitutionen ein Rückschluss auf die Messintention der Objektiv-Skalen gezogen werden konnte, wurden die hierzu verwendeten Items ebenfalls nach den Vorgaben für die Hybridklasse erstellt. Mit anderen Worten: Es wurde eine Maskierung der tatsächlichen Zieldimension angestrebt. Ähnliches galt für die Erfassung der Eingebundenheit in informelle und formelle Netzwerke. Das Vertrauen in Implementationsinstitutionen und die Eingebundenheit in informelle Netzwerke wurde anhand von Statements bewerkstelligt, die jeweils auf einer fünfstufigen Ratingskala bewertet werden sollten, deren Randausprägungen mit „*Stimme nicht zu*“ (1) und „*Stimme zu*“ (5) benannt waren. In den Itembatterien waren jeweils insgesamt vier sowohl negativ als auch positiv gepolte Items enthalten. Im Zuge der Datenauswertung wurde dies dementsprechend umgepolzt, sodass ein hoher Skalenwert einem hohen Institutionenvertrauen bzw. einer starken Eingebundenheit in informelle Netzwerke entsprach.

Die Eingebundenheit in formelle Netzwerke wurde anhand der Mitgliedschaft in verschiedenen Typen dieser Kategorie abgefragt. Hierbei waren Mehrfachnennungen erlaubt,

7. Hauptstudie

sodass eine hohe Anzahl von Mitgliedschaften ein starkes Engagement in formellen Netzwerken anzeigte.

7.1.1.4.7 Standard-Skala

Die hier als Standard-Skala bezeichnete Vertrauensskala stimmt im Wortlaut mit der deutschen Übersetzung der *Trust in People-Scale*, dem Standardinstrument zur Erfassung Generalisierten Vertrauens, überein. (Kunz (2004), S. 204f) Wie bereits an anderer Stelle⁶⁵ erwähnt, deckt die *Trust in People-Scale* insgesamt drei Dimensionen ab, die mit Vertrauen, Fairness und Hilfsbereitschaft beschrieben werden. Unterschiede zur *Trust in People-Scale* ergeben sich ausschließlich aus der angewandten Antwortskala. So verwendet die *Trust in People-Scale* standardmäßig pro Item nur zwei Antwortalternativen, zwischen denen die Probanden wählen können. Für die Vertrauenskomponente sind dies „*Can't be too careful*“ (0) und „*Most people can be trusted*“ (1), für die Hilfsbereitschaftskomponente „*Look out for themselves*“ (0) und „*Try to be helpful*“ und für die Fairnesskomponente „*Take advantage*“ (0) und „*Try to be fair*“ (1). Im Rahmen des Surveyexperimentes wird stattdessen je Skalenkomponente eine fünfstufige Ratingskala zur Bewertung der Statements eingesetzt, wobei die obengenannten Antwortmöglichkeiten als Randausprägungen bestehen bleiben. Diese Vorgehensweise bietet zwei Vorteile. So lässt sich mit Hilfe einer mehrstufigen Antwortskala mehr Varianz abbilden als anhand dichotomer Antwortalternativen. Es wird den Probanden ermöglicht, ein differenzierteres Urteil hinsichtlich ihres Vertrauenslevels abzugeben, und Validität sowie Reliabilität steigen. (Bühner (2011), S. 111) Zusätzlich bedeutet die Verwendung einer fünfstufigen Ratingskala eine Angleichung an das Antwortformat der Objektiv-Skalen, was eine Verbesserung der Vergleichbarkeit bewirkt.

Tab. 7.6: Items der Standard-Skala

Vertrauenskomponente

1. Ganz allgemein gesprochen: Glauben Sie, dass man den meisten Menschen vertrauen kann, oder dass man im Umgang mit anderen Menschen nicht vorsichtig genug sein kann?

Man kann nicht vorsichtig genug sein (1) (2) (3) (4) *Man kann vertrauen (5)*

Hilfsbereitschaftskomponente

2. Und glauben Sie, dass die Menschen versuchen hilfsbereit zu sein, oder dass die Menschen meistens auf den eigenen Vorteil bedacht sind?

Sind auf den eigenen Vorteil bedacht (1) (2) (3) (4) *Versuchen hilfsbereit zu sein (5)*

Fairnesskomponente

3. Glauben Sie, dass die meisten Menschen versuchen Sie auszunutzen, wenn Sie die Gelegenheit dazu haben, oder versuchen die meisten Menschen, sich fair zu verhalten?

Versuchen einen auszunutzen (1) (2) (3) (4) *Versuchen sich fair zu verhalten (5)*

(Quelle: Kunz (2004), S. 204f; Robinson/Shaver/Wrightsmann (1991), S. 406ff)

Um die kognitive Aufmerksamkeit der Probanden bei der Abfrage des Vertrauenslevels mit Hilfe der Standard-Skala auf einem möglichst hohen Niveau zu halten und Antwortmustern vorzubeugen, wurde im Fragebogen die Antwortskala der Fairnesskomponente invers zur Ausrichtung der Antwortskalen der verbleibenden Skalenkomponenten dargestellt.

⁶⁵ Vgl. Abschnitt „4.1 Standarditems zur Messung *generalisierten Vertrauens*“

7.1.1.4.8 Demografische Standards

Am Schluss des Fragebogens wurden die Probanden um Angaben hinsichtlich ihres Geschlechts, Alters und Studienfaches gebeten.

7.1.1.5 Auswertungskonzept

Wie bereits beschrieben⁶⁶, kann der für die Hauptstudie angefertigte Versuchsplan auch als Adaption des Solomon-Vier-Gruppen-Designs interpretiert werden. Entsprechend orientiert sich auch das Auswertungskonzept der Hauptstudie an der für diesen Versuchsplan erstellten Auswertungsstrategie.

Folgt man CAMPBELL/STANLEY, so lässt sich das Solomon-Vier-Gruppen-Design als 2x2-faktorieller-Versuchsplan begreifen, wobei sowohl das Treatment als auch der Pretest als Faktoren bzw. unabhängige Variablen angesehen werden. Es können dann drei Forschungshypothesen geprüft werden: Haupteffekt des Treatments⁶⁷, Haupteffekt des Pretests⁶⁸ und die Interaktion von Treatment und Pretest.⁶⁹ (Bortz/Döring (2016), S. 710ff; Braver/Braver (1988), S. 151; Campbell/Stanley (1969), S. 24f; Heller (2012), S. 187f)

Tab. 7.7: Analyseschema für Posttest-Ergebnisse des Solomon-Vier-Gruppen-Design

Pretest (B)	Treatment (A)	
	Ja	Nein
Ja	O ₂	O ₄
Nein	O ₃	O ₆

Die Schätzung des Haupteffektes des Treatments erfolgt dabei anhand der Differenz der Spaltenmittelwerte, während der Haupteffekt des Pretests durch die Differenz der Zeilenmittelwerte bestimmt wird. Unterscheiden sich die Spaltenmittelwerte signifikant voneinander, so spricht dies für einen Haupteffekt A. Unterscheiden sich die Zeilenmittelwerte signifikant voneinander, so zeigt dies einen Haupteffekt B an. Eine Interaktion von Treatment und Pretest ermittelt man anhand der zeilen- oder spaltenweisen Differenzen der Zellenwerte. Eine Interaktionseffekt AxB kann also dann vermutet werden, wenn sich die Zellenwerte zeilen- oder spaltenweise signifikant voneinander unterscheiden. (Bortz/Döring (2016), S. 710ff; Braver/Braver (1988), S. 151; Campbell/Stanley (1969), S. 24f; Huck/Sandler (1973))

Nach BORTZ/DÖRING erweitert sich das Kontinuum der innerhalb des Solomon-Vier-Gruppen-Designs wirksam werdenden Störvariablen um einen weiteren Effekt. So werden je nach betrachteter Gruppe drei unterschiedliche Arten von Effekten wirksam. Neben den bereits beschriebenen Effekten des Treatments (T), des Pretests (P) sind dies Effekte der Zeit (Z). Die Posttest-Ergebnisse der Gruppen stellen dabei jeweils eine Funktion dieser drei Effekte dar (Bortz/Döring (2016), S. 731):

$$PT_1 = f(T, P, Z)$$

⁶⁶ Vgl. Abschnitt „7.1.1.1 Erhebungskonzept“

⁶⁷ „Haupteffekt des Faktors A: Gibt es eine Wirkung des Faktors A, die über alle Stufen des Faktors B hinweg dieselbe ist?“ (Heller (2012), S. 188)

⁶⁸ „Haupteffekt des Faktors A: Gibt es eine Wirkung des Faktors B, die über alle Stufen des Faktors A hinweg dieselbe ist?“ (Heller (2012), S. 188)

⁶⁹ „Interaktionseffekt der beiden Faktoren A und B: Ist die Wirkung des Faktors A (bzw. B) unterschiedlich für verschiedene Stufen des Faktors B (bzw. A)?“ (Heller (2012), S. 188)

$$PT_2 = f(P, Z)$$

$$PT_3 = f(T, Z)$$

$$PT_4 = f(Z)$$

Wie aus dieser formalen Darstellung der Posttest-Ergebnisse hervorgeht, können Effekte der Zeit (Z) allerdings über alle Gruppen als konstant angenommen werden. Mit anderen Worten: Der Einfluss durch Zeiteffekte auf den Posttest ist in jeder Gruppe gleichartig gelagert und kann deshalb vernachlässigt werden. Um trotzdem eine Kontrolle von Zeiteffekten zu ermöglichen, müssen die Pretests in die Analyse miteinbezogen werden und die Annahme gültig sein, dass die Pretest-Ergebnisse nicht durch Zeiteffekte beeinflusst sind. Ergibt sich unter dieser Voraussetzung kein signifikanter Unterschied zwischen O_1 und O_6 bzw. O_3 und O_6 , so können Verzerrungen durch Zeiteffekte ausgeschlossen werden.

Unterschiede zwischen der Auswertungsstrategie des Solomon-Vier-Gruppen-Design und der Auswertungsstrategie des Surveyexperimentes im Rahmen der Hauptstudie sind vornehmlich darauf zurückzuführen, dass Effekte zeitgebundener Störvariablen durch die Art und Weise der Abfolge von Pretest und Posttest in *Gruppe 1: Objektiv/Standard* und *Gruppe 2: Standard/Objektiv* bereits ausgeschlossen wurden. So erfolgte die Präsentation der beiden Messmethoden zur Erfassung Generalisierten Vertrauens zwar sequenziell, allerdings wurde der zeitliche Abstand zwischen Pretest und Posttest minimiert bzw. der Posttest erfolgte unmittelbar im Anschluss an den Pretest. Unabhängig vom Effekt des Treatments (T_0 bzw. T_5 oder beides) können in den genannten Experimentalgruppen daher lediglich *Pretest-Effekte*, *Positions-Effekte* oder *Carry-over-Effekte* auftreten.

Ruft man sich die Definitionen dieser Effekte ins Gedächtnis, so wird deutlich, dass sie im Rahmen des Versuchsplans der Hauptstudie den gleichen Bias beschreiben, nämlich eine systematische Variation der abhängigen Variablen aufgrund des Pretests. Diese Schlussfolgerung ergibt sich unmittelbar aus der Tatsache, dass es sich beim Treatment des Surveyexperimentes um die Messmethode zur Erfassung Generalisierten Vertrauens bzw. deren Variation handelt. Die Probanden in *Gruppe 1: Objektiv/Standard* und *Gruppe 2: Standard/Objektiv* wurden somit insgesamt zwei Treatments ausgesetzt bzw. Pretest und Posttest entsprechen jeweils verschiedenen Versuchsbedingungen. Eine mögliche Beeinflussung der abhängigen Variablen durch den Pretest (*Pretest-Effekt*) entspricht demnach sowohl der Persistenz eines vorangegangenen Treatments (*Carry-over-Effekt*) wie auch Verzerrungen, die sich aus der Reihenfolge der experimentellen Bedingungen ergeben (*Positions-Effekt*). Insofern können *Pretest*-, *Positions*- und *Carry-over-Effekt* in der vorliegenden Studie als identischer bzw. äquivalenter Bias (B) betrachtet werden.

Für die Pretest- und Posttest-Ergebnisse des Surveyexperimentes ergeben sich daher folgende Funktionen:

Tab. 7.8: Funktionen der Pretest- und Posttest-Ergebnisse

Experimentalgruppe	Pretest	Posttest
<i>Gruppe 1: Objektiv/Standard</i>	$PR_1 (\text{Objektiv}) = f(T_0)$	$PT_1 (\text{Standard}) = f(T_5, B)$
<i>Gruppe 2: Standard/Objektiv</i>	$PR_2 (\text{Standard}) = f(T_5)$	$PT_2 (\text{Objektiv}) = f(T_0, B)$
<i>Gruppe 3: Objektiv</i>	---	$PT_3 (\text{Objektiv}) = f(T_0)$
<i>Gruppe 4: Standard</i>	---	$PT_4 (\text{Standard}) = f(T_5)$

Um die Effekte der verschiedenen Treatments (T_0/T_5) zu kontrastieren, werden Vergleiche der verschiedenen Messmethoden entlang des *within-subject*-Designs und entlang des *between-*

subject-Designs realisiert. Im Rahmen des *within-subject*-Designs lassen sich Unterschiede im Antwortverhalten der Probanden zwischen den Skalentypen durch einen Vergleich von Pretest und Posttest (PR_1 vs. PT_1 u. PR_2 vs. PT_2) umsetzen. Problematisch bei dieser Gegenüberstellung ist allerdings, dass nicht ohne Weiteres ausgeschlossen werden kann, ob der Unterschied zwischen beiden Messmethoden durch den oben beschriebenen Bias (Beeinflussung des Antwortverhaltens im Posttest durch den Pretest)⁷⁰ moderiert wurde. Vorteil des Vergleichs innerhalb der Gruppen, die einem *within-subject*-Design folgen, ist es, dass Verzerrungen, die auf spezifische Merkmale der Probanden zurückzuführen sind, zwischen Pretest und Posttest konstant gehalten werden. Kann eine Verzerrung durch den Pretest ausgeschlossen werden, so zeigt ein signifikantes Ergebnis, dass die Art und Weise der Messung Generalisierten Vertrauens Auswirkungen auf das Antwortverhalten der Probanden hat.

Da der Versuchsplan des Surveyexperimentes insgesamt vier voneinander unabhängige Gruppen umfasst, lässt sich ein möglicher Kontrast zwischen Objektiv-Skala und Standard-Skala auch anhand der Anordnung der Pretests und Posttests als *between-subject*-Designs untersuchen. Unterschiede im Antwortverhalten bzw. in der Ausprägung Generalisierten Vertrauens, die auf die Variation der Messmethode zurückzuführen sind, lassen sich dann anhand der folgenden Gegensätze isolieren: PT_3 vs. PT_4 ; PR_1 vs. PR_2 ; PR_1 vs. PT_4 und PR_2 vs. PT_3 . Generell gilt für dieses Arrangement der Pretests und Posttests, dass Verzerrungen, die sich aus einer Konfundierung der verschiedenen Treatments (T_0/T_5) ergeben können, eliminiert sind. Dieser Umstand ergibt sich daraus, dass jede der einzelnen Messungen als *One-Shot-Case-Study* bzw. *ex-post-facto-Design* interpretiert werden kann. (Bernard (2013), S. 104f) Gleichzeitig kann aber nicht ausgeschlossen werden, dass mögliche Unterschiede zwischen den Messmethoden auf Spezifika der Probanden innerhalb der jeweiligen Gruppen zurückzuführen sind. Ergeben sich für die genannten Gegenüberstellungen von Objektiv- und Standard-Skala signifikante Unterschiede, zeigt dies die Beeinflussung des Antwortverhaltens der Probanden durch die Variation der Messmethode an.

7.1.1.5.1 Kontrolltechniken

Die Kontrolle von *Pretest*-, *Positions*- und *Carry-over*-Effekten kann innerhalb der Hauptstudie auf prinzipiell drei Art und Weisen erfolgen.

Die erste Kontrollstrategie stellt die Methode des Lateinischen Quadrats dar. So wurden im Rahmen der *Versuchsgruppen 1* und *2* jeweils invers zueinander verlaufende Bearbeitungsreihenfolgen der Vertrauensskalen umgesetzt. Ergeben sich nun für PR_1 vs. PT_1 und PR_2 vs. PT_2 signifikante Unterschiede, die jeweils in die gleiche Richtung weisen, stellt dies einen Hinweis darauf dar, dass die Reihenfolge der Treatments nicht mit dem Effekt der Treatments konfundiert ist. Ergibt sich dagegen nur innerhalb einer der beiden Gruppen ein signifikanter Unterschied zwischen den beiden Messverfahren und/oder verläuft der ermittelte Unterschied in unterschiedliche Richtung je Gruppe, so muss ein *Positions*-Effekt angenommen werden. Wie bereits beschrieben, entsprechen sich *Pretest*-, *Positions*- und *Carry-over*-Effekt im Rahmen der Hauptstudie wechselseitig⁷¹, sodass bei einem positiven bzw. negativen Bescheid für einen dieser Effekte auch auf die Existenz bzw. Abwesenheit der übrigen Effekte geschlossen werden kann.

⁷⁰ Kontrollstrategien zur Aufdeckung einer solchen möglichen Verzerrungen werden im Abschnitt „7.1.1.5.1 Kontrolltechniken“ dargestellt.

⁷¹ Vgl. Abschnitt „7.1.1.5 Auswertungskonzept“

Die zweite Kontrollstrategie ergibt sich aus der Methode des Lateinischen Quadrats und beinhaltet einen Kreuzvergleich der Pretests und Posttests der *Gruppen 1* und *2*. So ermöglicht ein Vergleich von PR_1 und PT_2 bzw. PR_2 und PT_1 jeweils eine Abschätzung der obengenannten systematischen Verzerrungen⁷². Sowohl PR_1 als auch PT_2 erfassen Generalisiertes Vertrauen mit Hilfe der Objektiv-Skala (T_0), unterscheiden sich allerdings darin, dass das Resultat von PT_2 durch PR_2 verzerrt sein könnte, während PR_1 durch keine vorherige Messung belastet sein kann. Entsprechend gilt für PR_2 und PT_1 , dass beide Generalisiertes Vertrauen anhand der Standard-Skala (T_S) erfassen, wobei ausschließlich das Ergebnis von PT_1 durch PR_1 beeinflusst worden sein könnte, während PR_2 keinen Bias aufweist. Unterscheiden sich die Ergebnisse von PR_1 und PT_2 signifikant, so weist dies auf eine Beeinflussung des Antwortverhaltens auf der Objektiv-Skala durch die Standard-Skala in *Gruppe 2* hin. Unterscheiden sich die Ergebnisse von PR_2 und PT_1 signifikant, so weist dies auf eine Beeinflussung des Antwortverhaltens auf der Standard-Skala durch die Objektiv-Skala in *Gruppe 1* hin. Ergibt sich hingegen für keinen der Kreuzvergleiche ein signifikanter Unterschied, so kann angenommen werden, dass keine Verzerrung durch etwaige *Pretest*-, *Positions*- oder *Carry-over*-Effekte vorliegt.

Die zweite Kontrollstrategie ergibt sich aus der wechselseitigen Unabhängigkeit der *Experimentalgruppen 1 bis 4* und der daraus resultierenden Möglichkeit, diese in Form eines *between-subject*-Designs anzuordnen. So ermöglicht auch ein Vergleich von PT_1 und PT_4 sowie von PT_2 und PT_3 Rückschlüsse auf etwaige Verzerrungen, die auf den Einfluss des Pretests zurückzuführen sind. Sowohl PT_1 als auch PT_4 erfassen Generalisiertes Vertrauen mit Hilfe der Standard-Skala (T_S). Innerhalb dieser Paarung kann ausschließlich PT_1 durch Einflüsse eines vorangegangenen Treatments beeinflusst sein (PR_1), während im Vorhinein von PT_4 kein Pretest durchgeführt wurde. Ähnliches gilt für die Kontrastierung von PT_2 und PT_3 . Beide prüfen Generalisiertes Vertrauen anhand der Objektiv-Skala (T_0) ab, allerdings wurde nur im Fall von PT_2 ein Pretest (PR_2) durchgeführt, wohingegen PT_3 keine Effekte einer vorangegangenen Testung enthalten kann. Unterscheiden sich nun die Resultate von PT_1 und PT_4 signifikant, so weist dies auf eine Beeinflussung von PT_1 durch PR_1 hin. Unterscheiden sich die Ergebnisse von PT_2 und PT_3 signifikant, stellt dies einen Hinweis auf die Beeinflussung von PT_2 durch PR_2 dar. Geht aus keiner der beiden Kontrastierungen ein signifikantes Ergebnis hervor, so kann man davon ausgehen, dass in den *Gruppen 1* und *2* durch den jeweiligen Pretest keine Verzerrung des Posttest-Resultates bewirkt wird.

7.1.1.6 Validitätsbedrohungen und systematische Verzerrungen

Prinzipiell lassen sich im Rahmen von Experimenten Bedrohungen der internen und der externen Validität unterscheiden. Interne Validität ist dann gegeben, „wenn die beobachteten Änderungen in einer AV eindeutig (kausal) auf die Variation in der interessierenden UV zurückgeführt werden können“ (Bröder (2011), S. 49) bzw. wenn „Effekte in den abhängigen Variablen zweifelsfrei auf die Wirkung der unabhängigen Variablen zurückzuführen sind“ (Bortz/Döring (2016), S. 195). Dagegen beschreibt externe Validität die Generalisierbarkeit der Ergebnisse über die Untersuchungsbedingungen und den untersuchten Personenkreis hinaus.

⁷² Strenggenommen werden durch den Kreuzvergleich lediglich Positionseffekte aufgedeckt. Wie bereits in „7.1.1.5 Auswertungskonzept“ beschrieben, können im Rahmen der Hauptstudie *Positions*-, *Pretest*- und *Carry-over*-Effekte allerdings als identisch angenommen werden.

Da das Ziel der vorliegenden Untersuchung im Wesentlichen in der Bestimmung der kausalen Beziehung zwischen der Veränderung des Befragungsmodus und der daraus resultierenden Veränderung des Antwortverhaltens liegt und weniger die Generalisierung der Ergebnisse im Vordergrund steht, liegt der Fokus bei der Diskussion etwaiger Störvariablen auf Bedrohungen der internen Validität. Hinzu kommt, dass die externe Validität aufgrund der Verwendung eines *convenience sample* ohnehin als eingeschränkt angesehen werden kann.

Im Zuge der Darstellung des Erhebungskonzeptes wurden bereits zentrale Störfaktoren, wie *Pretest*-, *Positions*- und *Carry-over*-Effekte sowie Effekte der *demand characteristics*, diskutiert. Auf diese im Speziellen an das Design der Hauptstudie gebundenen Störvariablen soll nachfolgend nicht mehr eingegangen werden. Vielmehr stehen potentielle Konfundierungen im Mittelpunkt, die an die verwendeten Untersuchungsinstrumente gebunden sind bzw. sogenannte Item-Effekte.

7.1.1.6.1 Item-Nonresponse

Item-Nonresponse beschreibt die „explizite Verweigerung einer Antwort“ bzw. sogenannte Meinungslosigkeit („*Non-Attitudes*“) (Schnell/Hill/Esser (2008), S. 353). Um zu vermeiden, dass die Probanden bei einzelnen Items keine oder keine verwertbare Antwort geben, wurde im Vorhinein der Befragung zur korrekten Testbearbeitung aufgefordert. Im Speziellen wurde darauf hingewiesen, dass Antwortausfälle zu einer Eliminierung des gesamten Falles führen können und so vor dem Hintergrund, dass im Rahmen des Suveyexperimentes mit kleinen Fallzahlen gearbeitet wird, eine erhebliche Verzerrung der Ergebnisse aus dem Drop-out resultieren kann. Außerdem wurde auf die Schaltung einer expliziten „Weiß-nicht“-Kategorie verzichtet, um eine gültige Antwort zu erzwingen. (Schnell/Hill/Esser (2008), S. 337)

7.1.1.6.2 Fragereiheneffekte

Die Position eines Items innerhalb eines Fragebogens bzw. die Position eines Items relativ zu bestimmten anderen Items kann erheblichen Einfluss auf das Antwortverhalten der Probanden haben. Es werden dabei eine Vielzahl von Reihenfolgeeffekten unterschieden:

- (1) Ein nachfolgendes Item kann nicht unabhängig von einem vorangegangenen Item bearbeitet werden („*Aktualisierungseffekt*“). (Moosbrugger/Kelava (2012), S. 68) Die Problematik von Aktualisierungseffekten kann im Rahmen der Hauptstudie an zwei Stellen auftreten. So setzen sowohl die Bestimmung der Zieldimension der Objektiv-Skala als auch die Bestimmung des Vertrauensradius die vorherige Bearbeitung der jeweiligen Vertrauensskalen voraus. Irritationen seitens der Probanden wurde dadurch begegnet, dass im Rahmen der jeweiligen Items darauf hingewiesen wurde, dass die vorangegangene Itembatterie die Referenz für das abgefragte Urteil darstellen soll.
- (2) Die Probanden beantworten Items übereinstimmend, von denen sie glauben, dass sie das gleiche Merkmal erfassen („*Konsistenzeffekt*“). (Moosbrugger/Kelava (2012), S. 68) Konsistenzeffekte spielen insbesondere für die im *within-subject*-Design angeordneten Versuchsgruppen eine Rolle und korrespondieren hier mit den bereits beschriebenen *Pretest*-, *Positions*-, und *Carry-over*-Effekten. Sie können auftreten, wenn die Probanden aufgrund der Bearbeitung der Standard-Skala die Zieldimension der Objektiv-Skala antizipieren und dann je nach Präsentationsreihenfolge der Vertrauensskalen entweder prospektiv oder retrospektiv konsistent antworten. Mit anderen Worten: Aufgrund der

offenliegenden Zieldimension der Standard-Skala wird nun entweder das Antwortverhalten auf der Objektiv-Skala an das auf der Standard-Skala angepasst oder umgekehrt. Da die Probanden aufgrund des Erhebungsprozedere keine Änderungen mehr am Pretest vornehmen können, besteht in beiden Fällen ausschließlich die Möglichkeit einer retrospektiven Anpassung. Das bedeutet: In *Versuchsgruppe 1: Objektiv/Standard* besteht die Gefahr, dass die Probanden ihre Antworten auf der Standard-Skala an ihre Antworten auf der Objektiv-Skala anpassen, während in *Versuchsgruppe 2: Standard/Objektiv* die Gefahr besteht, dass die Probanden ihre Antworten auf der Objektiv-Skala an ihre Antworten auf der Standard-Skala anpassen. Da in *Versuchsgruppe 2* zuerst die Standard-Skala bearbeitet wird, ist die Chance für Konsistenzeffekte hier größer.

Darüber hinaus können Konsistenzeffekte auch auftreten, wenn die Probanden die im Rahmen der Objektiv-Skala angestrebte Verschleierung der Messintention durchschauen und deshalb auf Objektiv- und Standard-Skala übereinstimmend antworten. In der Konsequenz können sich aufgrund des Forschungsdesigns allerdings auch hierbei allenfalls retrospektive Konsistenzeffekte ergeben.

Die Eindämmung bzw. die Kontrolle von Konsistenzeffekten erfolgte durch die Verhinderung der Möglichkeit, Änderungen am Pretest vorzunehmen, und die Invertierung der Präsentationsreihenfolge von Objektiv- und Standard-Skala von *Versuchsgruppe 1* zu *Versuchsgruppe 2* bzw. umgekehrt.

- (3) Fragen am Ende eines Fragebogens werden weniger reflektiert beantwortet („*fatigue effect*“).
 - (4) Fragen am Anfang eines Fragebogens werden weniger reflektiert beantwortet („*rapport effect*“).
 - (5) Höherer Grad an Einverständnis mit Items am Anfang eines Fragebogens („*primacy effect*“) (Sudman/Bradburn/Schwarz (1996), S. 130)
 - (6) Höherer Grad an Einverständnis mit Items am Ende eines Fragebogens („*recency effect*“) (Sudman/Bradburn/Schwarz (1996), S. 130)
 - (7) Verschiedene Items überlappen in Hinsicht auf den abgefragten Inhalt, sodass der Eindruck entsteht, dass bestimmte Inhalte mehrfach abgefragt werden.
 - (8) Die Reihenfolge, in der die Items präsentiert werden, kann dazu führen, dass einzelne abgefragte Themengebiete an Stellenwert verlieren, weil verschiedene Kontexte aufeinandertreffen.
 - (9) Der *Halo*-Effekt stellt gewissermaßen den Prototypen eines Fragereiheneffektes dar und beschreibt ein „Ausstrahlen“ von einzelnen Items auf andere Items. Mit anderen Worten: Die Positionierung von Fragen innerhalb eines Fragebogens kann einen Einfluss auf das Antwortverhalten bei umgebenden Fragen haben. (Diekmann (2011), S. 464f)
- Ein Störeffekt dieser Art kann sich möglicherweise aus der Überschneidung der Items von Objektiv-Skala und Objektiv-Kurzskala ergeben. So bekommen die Probanden im Rahmen des Surveyexperimentes strenggenommen nur die Objektiv-Skala präsentiert. Die Herauslösung der zur Objektiv-Kurzskala zusammengefassten Items erfolgt dagegen erst im Zuge der Datenauswertung. Die Beantwortung der Items der Objektiv-Kurzskala ist somit nicht unabhängig von der Beantwortung der Items der Objektiv-Skala.

Eine potentielle Gegenmaßnahme hätte eine gesonderte Präsentation der Objektiv-Kurzskala darstellen können. Hierbei ergaben sich prinzipiell zwei Szenarien, die allerdings aus unterschiedlichen Gründen verworfen wurden. Zunächst hätte eine weitere Versuchsgruppe eröffnet werden können. Dies wurde allerdings ausgeschlagen, da der

organisatorische Aufwand zu groß war. Insbesondere die Rekrutierung weiterer Probanden gestaltete sich schwierig.

Eine weitere Möglichkeit, eine von der Objektiv-Skala separierte Präsentation der Objektiv-Kurzskala zu bewerkstelligen, bestand darin die Objektiv-Kurzskala im Vor- oder Nachhinein der Objektiv-Skala in den bereits bestehenden Fragebogen einzubetten. Die Probanden hätten dann allerdings den Eindruck gewonnen, dass bestimmte Inhalte mehrfach abgefragt würden, was möglicherweise weitere Verzerrungseffekte, wie beispielsweise einen Konsistenzeffekt nach sich gezogen hätte.

Keine der beiden hier beschriebenen Lösungsmöglichkeiten hätte allerdings die nach wie vor bestehende teilweise inhaltliche Äquivalenz und die daraus resultierende Ähnlichkeit der Messwerte der Objektiv-Skalen verhindert. Eine Unabhängigkeit des Antwortverhaltens auf den beiden Objektiv-Skalen hätte nur im Rahmen des ersten Szenarios gewährleistet werden können, was aber aus ökonomischen Gründen abgelehnt werden musste.

7.1.1.6.3 Antworttendenzen und Urteilsfehler

Antworttendenzen bzw. Urteilsfehler stellen eine Klasse systematischer Fehlerquellen dar, die sich infolge menschlicher Informationsverarbeitung (z.B. Urteilsheuristiken) ergeben oder mit Persönlichkeitsmerkmalen der Probanden verbunden sind. Sie beschreiben die Neigung zu bestimmtem Antwortverhalten abhängig von bestimmten individuellen Merkmalen der Versuchspersonen bzw. die Tendenz bei der Einstufung eigenen Erlebens und Verhaltens, von Aussagen oder Objekten die jeweilige Merkmalsausprägung mehr oder weniger stark zu über- bzw. zu unterschätzen. Dabei wird eine Reihe von möglichen Störeffekten unterschieden:

- (1) Im Allgemeinen beschreibt der Ausstrahlungseffekt („*Halo-Effekt*“) eine Tendenz zu globalen Pauschalurteilen. Dabei differenzieren die Probanden bei der Bewertung von Objekten nicht zwischen voneinander verschiedenen oder unabhängigen Merkmalen, vielmehr strahlt der Eindruck bezüglich eines Merkmals auf die Einstufung der übrigen Merkmale aus. Im Rahmen von Fragenbogenitems beschreibt der Halo-Effekt das Ausstrahlen einzelner Fragen auf andere Fragen und daraus resultierende Verzerrungen im Antwortverhalten⁷³ (Bortz/Döring (2016), S. 252f; Diekmann (2011), S. 464)

Zunächst kann ein *Halo-Effekt* aus der inhaltlichen Nähe der Items, die im Kontext der Objektiv- bzw. der Standard-Skala verwendet werden, resultieren, in deren Folge die Probanden auf den einzelnen Vertrauensskalen zu Pauschalurteilen neigen.

Eine Manifestation des *Halo-Effekts* könnte im Rahmen der Hauptstudie möglicherweise auch in enger Verbindung zu den *demand characteristics* der Objektiv-Skala stehen. Die Verschleierung der Messintention der Objektiv-Skala durch eine geringe Augenscheinvalidität der verwendeten Items und den Einsatz einer Coverstory bewirkt möglicherweise, dass dem Antwortverhalten der Probanden eine subjektive Interpretation der Zieldimension zugrunde liegt, die von der tatsächlichen Zieldimension abweicht. Die Bewertung der einzelnen Items könnte dann von einem spezifischen Marker-Item abhängig sein, aus dem die Probanden die potentielle Zieldimension abgeleitet haben.

- (2) Der Milde-Härte-Fehler („*Leniency-Severity-Fehler*“) ist dadurch definiert, dass Objekte oder Personen systematisch zu positiv oder zu negativ bewertet werden. (Bortz/Döring (2016), S. 253) Fehlurteile dieser Art sind im Rahmen der Hauptstudie voraussichtlich eng

⁷³ In diesem Sinne stellt der *Halo-Effekt* einen sogenannten *Fragereiheneffekt* dar. (Diekmann (2011), S. 464)

verknüpft mit dem Radiusproblem. So besteht die Möglichkeit, dass entweder ein an konkrete Andere gebundenes Vertrauen besonders hoch ausfällt, während ein an generalisierte Andere gebundenes Vertrauen besonders niedrig ausfällt oder umgekehrt. Abhängig davon, welche Referenzgruppe die Probanden jeweils angegeben haben, würde dies eine Verzerrung des wahren Ausmaßes an Vertrauen nach sich ziehen.

- (3) Die Zustimmungstendenz („*Akquieszenz*“) beschreibt die Neigung der Probanden, unabhängig vom Inhalt einer Frage eine zustimmende Antwort zu geben. (Bortz/Döring (2016), S. 255; Bühner (2011), S. 129; Mummendey/Grau (2014), S. 156ff; Schnell/Hill/Esler (2008), S. 354f)

Antworttendenzen dieser Art können prinzipiell bei allen Items des eingesetzten Untersuchungsinstrumentes auftreten und sollten sich in Item übergreifend geringen Antwortvarianzen bzw. der Zustimmung zu invers zueinander formulierten Statements manifestieren. Insbesondere lange Itembatterien oder solche, die unterschiedlich gepolte Items enthalten, erscheinen prädestiniert für solche Störeffekte. Als Gründe hierfür lassen sich ein Absinken der kognitiven Aufmerksamkeit oder Desinteresse am Inhalt der Befragung vermuten. (Bühner (2011), S. 129; Häcker/Schwenkmezger/Utz (1979), S. 15f; Moosbrugger/Kelava (2012), S. 61)

- (4) Mit Tendenz zur Mitte („*Zentrale Tendenz*“) wird ein Urteilsfehler beschrieben, bei dem die Probanden dazu neigen, grundsätzlich moderate Bewertungen vorzunehmen und extreme Antworten zu vermeiden. (Bortz/Döring (2016), S. 253f; Korman (1971), S. 180f)

Eine einfache Variante zur Verhinderung einer Tendenz zur Mitte ist die Vermeidung einer neutralen Mittelkategorie innerhalb der verwendeten Antwortskalen und das Angebot einer „*Weiß-nicht*“-Kategorie. (Moosbrugger/Kelava (2012), S. 60f) Auf beide Maßnahmen wurde im Rahmen der Hauptstudie allerdings verzichtet. So erschien der Verzicht auf eine neutrale Mittelkategorie als Einschränkung für die Beurteilungen der Items durch die Probanden und wurde deshalb verworfen. So hätten Antwortskalen mit einer geraden Anzahl von Antwortalternativen möglicherweise eine künstlich herbeigeführte Polarisierung des Vertrauenslevels bewirkt. Auf eine „*Weiß-nicht*“-Kategorie wurde verzichtet, um die Gefahr von *Missing values* vor dem Hintergrund der niedrigen Fallzahlen in den verschiedenen Versuchsgruppen des Surveyexperimentes möglichst gering zu halten.

7.1.1.6.4 Verfälschbarkeit

Die Tendenz zur Verfälschung von Testwerten kann im Wesentlichen zwei verschiedene Formen annehmen:

- (1) Simulation („*fake good*“): Vortäuschen von Verhalten, das man sonst nicht zeigt.
- (2) Dissimulation („*fake bad*“): Verschleiern oder Verbergen von Verhalten, das man normalerweise zeigt.

Im Allgemeinen beschreibt die Verfälschungstendenz somit ein Antwortverhalten, das sich an taktischen Überlegungen orientiert, anstatt von Ehrlichkeit geprägt zu sein. Sie tritt insbesondere infolge von Items auf, die eine Selbstauskunft oder Selbstbeurteilung erfordern. (Bortz/Döring (2016), S. 436f; Bühner (2011), S. 125ff)

Das bekannteste Konzept stellt in diesem Zusammenhang sozialerwünschtes Antwortverhalten dar. Soziale Erwünschtheit („*Social Desirability*“) beschreibt im Wesentlichen eine Tendenz der Probanden sich in ihrem Antwortverhalten weniger an ihrem tatsächlichen

7. Hauptstudie

Verhalten als an sozialen Normen zu orientieren. Sozialerwünschtes Antwortverhalten tritt dabei vornehmlich in Situationen auf, in denen man sich besonders gut darstellen möchte, in denen negative Konsequenzen aus einem nicht-normkonformen Verhalten drohen oder in denen man ein positives Bild von sich erhalten möchte. Dabei sollen positives Verhalten bzw. sozial konforme Verhaltensweisen oder besonders günstige Eigenschaften oder Merkmale hervorgehoben werden, während gleichzeitig als negativ empfundenes Verhalten bzw. normverletzende Verhaltensweisen oder unerwünschte Eigenschaften und Merkmale verborgen werden sollen. Welches Verhalten dabei als positiv oder negativ gilt, ist dabei individuell, je nach Gruppe und situationsabhängig verschieden. (Bortz/Döring (2016), S. 437ff; Bühner (2006), S. 60ff; Schnell/Hill/Esser (2008), S. 355f) Es existiert eine Vielzahl von Verfahren, um den Einfluss sozialer Erwünschtheit auf das Antwortverhalten in Fragebögen zu kontrollieren (Bortz/Döring (2016), S. 438f; Bühner (2011), S. 127f):

- (1) *Aufforderung zu korrekter Testbearbeitung*: Der Befragung werden Testinstruktionen vorangestellt, die zu ehrlichen Antworten aufrufen.
- (2) *Ausbalancierte Antwortvorgaben*: Alle Antwortalternativen eines Items besitzen denselben Grad an sozialer Erwünschtheit.
- (3) *Forced-Choice-Items*: Dem Probanden werden verschiedene Aussagen vorgelegt, wobei mindestens eine dieser Aussagen als Antwort ausgewählt werden muss.
- (4) *Random-Response-Technik*: Die geprüfte Person kann sich aufgrund eines speziellen Zufallsverfahrens absolut sicher sein, dass sich ihr Antwortverhalten nicht rekonstruieren lässt. Für jedes Item wird per Zufall entschieden (z.B. Würfeln), ob ehrlich geantwortet wird. Es wird nun angenommen, dass aufgrund der geringen Rückverfolgungschancen die Probanden nun keine Veranlassung mehr zur Verfälschung ihrer Antworten haben.
- (5) *Einsatz von Kontrollskalen*: Es werden Items bzw. sogenannte SD-Skalen in den Fragebogen aufgenommen, die besonders sensibel auf sozialerwünschtes Antwortverhalten reagieren.
- (6) *Objektive Tests*: Die Zieldimension bzw. das Testziel wird durch geschickte Aufgabenauswahl und Auswertungstechnik für die Probanden möglichst undurchschaubar gestaltet.
- (7) *Bogus-Pipeline-Methode*: Während der Bearbeitung des Fragebogens sind die Probanden an einen vermeintlichen Lügendetektor angeschlossen, was in einer signifikanten Reduktion von Falschantworten resultiert.
- (8) *Faking-Instruktionen*⁷⁴: Das Testinstrument wird von den Probanden unter der Bedingung, sich möglichst positiv darzustellen, sich möglichst negativ darzustellen, und unter normalen Bedingungen bearbeitet. Aus dem Vergleich des jeweiligen Antwortverhaltens lassen sich Rückschlüsse auf die Anfälligkeit des Untersuchungsinstrumentes für Verfälschungen ableiten.

Zur Kontrolle bzw. Erfassung sozialerwünschten Antwortverhaltens kommt im Rahmen der Hauptstudie eine Kontrollskala zum Einsatz. Das zu diesem Zweck verwendete *Balanced Inventory of Desirable Responding* (BIDR) gliedert soziale Erwünschtheit dabei in zwei Dimensionen: Selbsttäuschung (*self-deceptive enhancement*, SDE) und Fremdtäuschung (*impression management*, IM). Die Verwendbarkeit Objektiver Tests zur Reduktion sozialerwünschten Antwortverhaltens ist Gegenstand der hier vollzogenen empirischen Untersuchung.

⁷⁴ Eine Diskussion des Einsatzes von *Faking-Instruktionen* im Rahmen des Forschungsdesigns der Hauptstudie ist Gegenstand des Abschnittes „7.1.1.7 Alternatives Forschungsdesign“.

7.1.1.6.5 Hawthorne-Effekt

Verhaltensänderungen der Probanden, die auf das Bewusstsein Gegenstand einer wissenschaftlichen Untersuchung bzw. einer Testsituation zurückzuführen sind, unabhängig von der Wirksamkeit eines Treatments, werden als *Hawthorne-Effekt* bezeichnet. (Bortz/Döring (2016), S. 101; Diekmann (2011), S. 341; Eid/Gollwitzer/Schmitt (2015), S. 53)

Veränderungen im Antwortverhalten der Probanden aufgrund eines Hawthorne-Effektes können möglicherweise in den Versuchsgruppen, die einem experimentellen Design mit Messwiederholung folgen (*within-subject-Design*), eine Rolle spielen. Überlegungen seitens der Probanden, welche die inhaltliche Verbindung zwischen Pretest und Posttest betreffen, könnten Verzerrungen in den Messergebnissen des jeweiligen Posttests bewirken.

7.1.1.7 Alternatives Forschungsdesign

Das für die Hauptstudie gewählte Forschungsdesign stellt nur eine Möglichkeit der Erhebungsgestaltung dar, weshalb nachfolgend auch die Vor- und Nachteile eines alternativen Forschungsdesigns diskutiert werden sollen. Ausgangspunkt der Argumentation ist dabei die Erfassung der Neigung der Probanden zu sozialerwünschtem Antwortverhalten bzw. die Messung des Einflusses sozialer Erwünschtheit auf die Resultate der verschiedenen Messverfahren.

So könnte ein mögliches alternatives Forschungsdesign sogenannte *faking*-Instruktionen zur Ermittlung der Anfälligkeit der eingesetzten Vertrauensskalen für sozialerwünschtes Antwortverhalten verwenden. Hierbei werden mehrere Versuchsgruppen gebildet, innerhalb derer die Probanden die Skalen unter *faking-good*-, *faking-bad*- oder Normalbedingungen bearbeiten sollen. Die *faking-good*-Instruktion hält die Probanden dazu an, sich bezüglich der Zieldimension der Skala möglichst positiv darzustellen, während die *faking-bad*-Instruktion vorgibt, einen möglichst negativen Eindruck zu erzeugen. Unter Normalbedingungen erhalten die Probanden keine besondere Verhaltensaufforderung, sondern bearbeiten die Skalen nach eigenem Ermessen. Je größer die Differenz zwischen dem Messergebnis der Normalbedingung und den Messergebnissen der *faking*-Bedingungen ausfällt, desto größer ist die Anfälligkeit der Skalen für Verfälschungen. (Bortz/Döring (2016), S. 437; Mummendey (1981), S. 207f; Scholl (2015), S. 221)

Vorteil einer solchen Vorgehensweise ist, dass die Messung sozialerwünschten Antwortverhaltens unabhängig von einer *ex-ante*-Setzung dessen erfolgt, was als sozialerwünscht angenommen wird. Mit anderen Worten: Man lässt die Probanden selbstbestimmen, was in Bezug auf die Zieldimension als Verfälschung des wahren Wertes gilt. Auf diese Weise lassen sich einige Probleme, die aus der Messung der Tendenz zu sozialerwünschtem Antwortverhalten mittels sogenannter SD-Skalen resultieren, umgehen. Solche Kontrollskalen setzen sich aus Items zusammen, die „typischerweise Eigenschaften oder Verhaltensweisen, die allgemein negativ beurteilt werden, aber doch so oft vorkommen, dass eine ablehnende Antwort unglaublich erscheint“ (Bortz/Döring (2016), S. 439), erfassen. Eine hohe Korrelation zwischen SD-Skala und interessierendem Test deutet dann auf eine Verzerrung der Testwerte in Richtung sozialer Erwünschtheit hin.

Aus der Anwendung von SD-Skalen können allerdings verschiedene Problematiken entstehen. So kann es vorkommen, dass die Probanden die fragliche Eigenschaft oder Verhaltensweise tatsächlich nicht besitzen oder zeigen, sodass eine ablehnende Antwort keinem

sozialerwünschten Antwortverhalten entspricht. Hierbei spielt möglicherweise aber auch eine Rolle, dass die Probanden im Alltag nach sozialer Anerkennung streben und deshalb grundsätzlich sozialerwünscht handeln, sodass solches Verhalten ihrer natürlichen Handlungsweise entspricht. (Scholl (2015), S. 221) Unabhängig davon würde eine solche Antwort aufgrund des Analyseschemas der SD-Skala trotzdem als sozialerwünscht gewertet werden. Mit anderen Worten: Der Einsatz von Kontrollskalen ermöglicht Fehlinterpretationen, die in einer Überschätzung der Tendenz zu sozialerwünschtem Antwortverhalten resultieren können. Zudem gelten auch SD-Skalen als nicht verfälschungssicher. Gelingt es den Probanden die Messintention einer Kontrollskala zu durchschauen, kann dies dazu führen, dass systematisch SD-positive Items abgelehnt werden und SD-negativen Items zugestimmt wird. In der Folge würde eine Tendenz zu sozialerwünschtem Verhalten nicht mehr hohen Werten, sondern niedrigen Werten auf der SD-Skala entsprechen. (Herzberg (2011), S. 126f) Eine per SD-Skala ermittelte Tendenz zu sozialerwünschtem Antwortverhalten muss sich außerdem nicht unbedingt auf das übrige Antwortverhalten innerhalb eines Tests übertragen. (Scholl (2015), S. 221)

Auf die Umsetzung des hier beschriebenen alternativen Forschungsdesigns wurde im Rahmen der Hauptstudie vornehmlich aus ökonomischen und methodischen Gründen verzichtet. Um weiterhin die Kontrolle von Störvariablen, die sich aus der experimentellen Anordnung ergeben (*Pretest*-, *Positions*- und *Carry-over*-Effekt), zu ermöglichen, wäre eine Aufstockung auf insgesamt 12 Versuchsgruppen notwendig gewesen. So hätten innerhalb jeder der bereits bestehenden vier Versuchsgruppen Subgruppen entsprechend der drei unterschiedlichen *faking*-Instruktionen gebildet werden müssen (4x3 Gruppen=12 Gruppen). Dies hätte entweder zur Folge gehabt, dass die Fallzahl für jede der *faking*-Instruktionen in den einzelnen Versuchsgruppen stark limitiert ist oder es hätte eine Aufstockung der Fallzahl erfolgen müssen. Die geringe Fallzahl hätte eine Einschränkung der Validität der Ergebnisse bedeutet, während die Akquise weiterer Studienteilnehmer organisatorisch kaum möglich war.

Zudem ermöglichte die im Kontext der Hauptstudie verwendete BIDR-Kurzskala eine trennscharfe Messung der verschiedenen Komponenten sozialer Erwünschtheit. Dieser Umstand war für die nachfolgende Datenanalyse von besonderer Relevanz.

Wichtigster Entscheidungsgrund für die Verwendung von SD-Skalen war allerdings, dass diese eine von der Zieldimension unabhängige Messung der Tendenz zu sozialerwünschtem Antwortverhalten erlauben. Eine der Kerncharakteristika der Objektiv-Skala war die geringe Augenscheinvalidität der in ihr zusammengefassten Items. Die Nutzung des Alternativdesigns hätte allerdings gegebenenfalls notwendig gemacht, den Probanden die Messintention der Skala offenzulegen, da ansonsten keine zielgerichtete Befolgung der *faking*-Instruktionen möglich gewesen wäre. Auf diese Weise wäre die Objektiv-Skala jedoch ihrer potentiellen Fähigkeit zur Vermeidung von bewussten Antwortmodifikationen beraubt worden.

7.2 Datenauswertung

Die Ergebnisse der Datenauswertung basieren auf Analysen mit Hilfe der statistischen Analyseprogramme *SPSS* und *G*Power*. Letzteres wird im Rahmen der Mittelwertvergleiche zur Bestimmung von Teststärke und β -Fehler-Wahrscheinlichkeit sowie der zugehörigen grafischen Darstellungen eingesetzt. (Faul/Erdfelder/Lang/Buchner (2007); Faul/Erdfelder/Lang/Buchner (2009))

7.2.1 Deskription

Im Rahmen dieses Abschnittes erfolgt eine deskriptive Analyse ausgesuchter Variablen des innerhalb der Hauptstudie eingesetzten Fragebogens.

7.2.1.1 Aktuelle und allgemeine Lebenszufriedenheit

Wie bereits beschrieben⁷⁵, wird ein positiver Zusammenhang zwischen aktueller und allgemeiner Lebenszufriedenheit und der Ausprägung Generalisierten Vertrauens angenommen. Personen, die mit ihrem Leben zufrieden sind, zeichnen sich durch eine tendenziell optimistischere Weltwahrnehmung aus, die wiederum eine wohlwollende Perspektive auf Menschen im Allgemeinen bewirkt und sich ebenfalls vorteilhaft auf den Grad Generalisierten Vertrauens auswirkt.

Tab. 7.9: Häufigkeitsverteilung: Aktuelle und allgemeine Lebenszufriedenheit; Gruppe 1: Objektiv/Standard (N=50)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Zufriedenheit/Glück aktuell	4,02	,820	-,962	1,029	2	5
Lebenszufriedenheit allgemein	3,94	,818	-,818	,675	2	5

^a Arithmetisches Mittel

^b Standardabweichung

^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)

^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Aus den Mittelwerten geht hervor, dass sich weder die aktuelle noch die allgemeine Lebenszufriedenheit zwischen den Gruppen nennenswert unterscheiden. Auch lässt sich beobachten, dass die Differenz zwischen aktueller und allgemeiner Lebenszufriedenheit eher gering ausfällt. Insgesamt sind die Probanden auf beiden Ebenen eher zufrieden mit ihrem Leben.

Tab. 7.10: Häufigkeitsverteilung: Aktuelle und allgemeine Lebenszufriedenheit; Gruppe 2: Standard/Objektiv (N=50)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Zufriedenheit/Glück aktuell	3,76	,894	-1,464	5,245	0	5
Lebenszufriedenheit allgemein	4,04	,781	-,873	1,067	2	5

^a Arithmetisches Mittel

^b Standardabweichung

^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)

^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Auch die Streuung der Lebenszufriedenheit bleibt über die Gruppengrenzen hinweg stabil. Dabei variiert die tagesaktuelle Lebenszufriedenheit etwas stärker als die allgemeine Lebenszufriedenheit. Dies war zu erwarten, da auf die aktuelle Lebenszufriedenheit insbesondere kurzfristige Erlebnisse und Ereignisse Einfluss ausüben.

⁷⁵ Vgl. „Abschnitt 6.1.1.4.1 Glück und subjektives Wohlbefinden/Zufriedenheit“

Tab. 7.11: Häufigkeitsverteilung: Aktuelle und allgemeine Lebenszufriedenheit; Gruppe 3: Objektiv (N=52)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Zufriedenheit/Glück aktuell	3,63	,817	-,565	-,085	2	5
Lebenszufriedenheit allgemein	3,96	,791	-,672	,486	2	5

^a Arithmetisches Mittel^b Standardabweichung^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Gruppenübergreifend stellen die Häufigkeitsauszählungen beider Variablen rechtssteile Verteilungen dar. Dies unterstützt den durch die Mittelwerte erlangten Eindruck, dass die Probanden prinzipiell zu einer hohen Lebenszufriedenheit neigen.

Tab. 7.12: Häufigkeitsverteilung: Aktuelle und allgemeine Lebenszufriedenheit; Gruppe 4: Standard (N=52)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Zufriedenheit/Glück aktuell	3,54	,896	-,971	1,137	1	5
Lebenszufriedenheit allgemein	3,92	,682	-,290	,267	2	5

^a Arithmetisches Mittel^b Standardabweichung^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Mit einer Ausnahme in *Versuchsgruppe 3* ergeben sich für die beiden Variablen ausschließlich schmalgipflige Verteilungen. Dies unterstützt nochmals den Befund, dass sich die Antworten der Probanden auf der rechten Seite der Skala sammeln und ein hohes Maß an Lebenszufriedenheit vorherrscht.

Insgesamt scheinen die Probanden über die Gruppengrenzen hinweg eine, was die Lebenszufriedenheit angeht, homogene Schicht darzustellen.

7.2.2 Aggregation

Für die Datenauswertung wurden verschiedene Itembatterien zu additiven Indizes zusammengefasst.

7.2.2.1 Vertrauensskalen

Generalisiertes Vertrauen stellt eine latente Variable dar, deren Ausprägung sich im Summenwert der Itemantworten der Objektiv-Skala bzw. Objektiv-Kurzskala und der Standard-Skala manifestieren. Die Berechnung dieses Summenwertes erfolgt dabei gruppenweise, da jede Versuchsgruppe eine von den anderen Versuchsgruppen unabhängige Messung darstellt. In die nachfolgenden Analysen gehen ausschließlich die Summenscores der Vertrauensskalen bzw. die gemittelten Summenscores der Vertrauensskalen ein.

Tab. 7.13: Summenscore: Generalisiertes Vertrauen; Gruppe 1: Objektiv/Standard (N=50)

	\bar{x}_i^d	σ_i^e	Sch ^f	Ex ^g	Min	Max
PR ₁ ^a	2,88	,52060	-,171	,670	2	4
PR _{1(KS)} ^b	3,00	,63888	,000	-,424	2	4
PT ₁ ^c	3,28	,75701	-,230	-,683	2	5

^a Objektiv-Skala^b Objektiv-Kurzskala^c Standard-Skala^d Arithmetisches Mittel^e Standardabweichung^f Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)^g Exzess bzw. Kurtosis (Ex_i < 0 → breitgipflige Verteilung; Ex_i > 0 → schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Vergleicht man die Mittelwerte der Summenscores der verschiedenen Vertrauensskalen über alle Experimentalgruppen hinweg, so fällt auf, dass auf den Objektiv-Skalen durchweg ein niedrigeres durchschnittliches Generalisiertes Vertrauen angegeben wird als auf der Standard-Skala. Dies stellt einen ersten Hinweis daraufhin dar, dass die Probanden möglicherweise auf der Standard-Skala ihr Vertrauenslevel überschätzen. Es ist allerdings zu bemerken, dass der Unterschied zwischen Objektiv- und Standard-Skala gruppenübergreifend maximal 0,4 Skalenpunkte und zwischen Objektiv-Kurzskala und Standard-Skala gruppenübergreifend maximal 0,34 Skalenpunkte beträgt. Die Differenz fällt somit vergleichsweise gering aus. Es ist die Aufgabe nachfolgender Signifikanzprüfungen, ob es sich hierbei um überzufällige Unterschiede handelt.

Tab. 7.14: Summenscore: Generalisiertes Vertrauen; Gruppe 2: Standard/Objektiv (N=50)

	\bar{x}_i^d	σ_i^e	Sch ^f	Ex ^g	Min	Max
PR ₂ ^a	3,34	,79821	,050	-,405	2	5
PT ₂ ^b	2,96	,49322	-,100	1,405	2	4
PT _{2(KS)} ^c	3,14	,53490	,135	,410	2	4

^a Standard-Skala^b Objektiv-Skala^c Objektiv-Kurzskala^d Arithmetisches Mittel^e Standardabweichung^f Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)^g Exzess bzw. Kurtosis (Ex_i < 0 → breitgipflige Verteilung; Ex_i > 0 → schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Betrachtet man die Standardabweichungen der Summenscores, fällt auf, dass die Standard-Skala eine leicht differenziertere Messung Generalisierten Vertrauens bereitstellt als die Objektiv-Skalen. Dieser Eindruck wird auch dadurch unterstützt, dass die Standard-Skala im Vergleich mit den Objektiv-Skalen ein etwas breiteres Spektrum der Antwortskala nutzt. So ist der auf allen Skalen erreichte Minimalwert 2, wobei der Maximalwert der Standard-Skala bei 5 liegt und der Maximalwert der Objektiv-Skalen bei jeweils 4. Allerdings ist auch hier zu bemerken, dass der beobachtete Unterschied verhältnismäßig klein ausfällt.

Tab. 7.15: Summenscore: Generalisiertes Vertrauen; Gruppe 3: Objektiv (N=52)

	\bar{x}_i^c	σ_i^d	Sch ^e	Ex ^f	Min	Max
PT ₃ ^a	3,0192	,61006	-,009	-,165	2	4
PT _{3(KS)} ^b	3,0769	,58899	-,011	,005	2	4

^a Objektiv-Skala^b Objektiv-Kurzskala^c Arithmetisches Mittel^d Standardabweichung^e Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)^f Exzess bzw. Kurtosis (Ex_i < 0 → breitgipflige Verteilung; Ex_i > 0 → schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Was die Schiefe der Verteilungen der Summenscores angeht, so lässt sich weder für die Objektiv-Skalen noch für die Standard-Skalen eine eindeutige gruppenübergreifende Tendenz feststellen, was die Ausrichtung der Schiefe angeht. Auffällig ist allenfalls, dass in den Versuchsgruppen, die als *within-subject*-Design konzipiert wurden, die Verteilung des jeweiligen Posttests eine größere Schiefe aufweist, als die Verteilung des Pretests. Dies deutet daraufhin, dass hier möglicherweise Interaktionseffekte vorliegen bzw. die Bearbeitung des Posttests nicht unabhängig von der Bearbeitung des Pretests erfolgte. Außerdem zeigt sich, dass wenn die Vertrauensskalen voneinander getrennt präsentiert werden, wie in den *Experimentalgruppen 3* und *4* geschehen, die Objektiv-Skalen nahezu Symmetrie aufweisen, während die Standard-Skala leicht rechtssteil ausfällt. Auf der Standard-Skala neigen die Probanden somit tendenziell dazu, ein höheres Vertrauenslevel anzugeben.

Tab. 7.16: Summenscore: Generalisiertes Vertrauen; Gruppe 4: Standard (N=52)

	\bar{x}_i^b	σ_i^c	Sch _i ^d	Ex _i ^e	Min	Max
PT ₄ ^a	3,1923	,81742	-,150	-1,037	2	5

^a Standard-Skala^b Arithmetisches Mittel^c Standardabweichung^d Schiefe (Sch_i < 0 → rechtssteile Verteilung; Sch_i > 0 → linkssteile Verteilung) (Bühner (2006), S. 88)^e Exzess bzw. Kurtosis (Ex_i < 0 → breitgipflige Verteilung; Ex_i > 0 → schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Auch hinsichtlich der Ausprägung des Exzess bzw. der Kurtosis lässt sich keine allgemeine Tendenz für die Objektiv-Skalen und die Standard-Skala über alle Gruppen hinweg finden. Es zeigt sich allerdings, dass in den als *within-subject*-Design angelegten Experimentalgruppen der Summenscore der Objektiv-Skala jeweils zu einer schmalgipfligen Verteilung neigt, während die Standard-Skala jeweils zu einer breitgipfligen Verteilung tendiert.

7.2.2.2 Korrelate Generalisierten Vertrauens

Wie bereits beschrieben⁷⁶, wurden zur Bestimmung der Kriteriumsvalidität der Objektiv-Skalen sogenannte externe Kriterien bzw. Dimensionen, die im Zusammenhang mit Generalisiertem Vertrauen stehen, erhoben. Im Einzelnen waren dies das Vertrauen in Implementationsinstitutionen und die Eingebundenheit in informelle bzw. formelle Netzwerke. Jede dieser Variablen geht in die nachfolgende Validitätsanalyse als additiver Index ein. Auch für

⁷⁶ Vgl. Abschnitt „5.3 Validität“

7. Hauptstudie

diese deskriptive Analyse gilt, dass jede Experimentalgruppe eine unabhängige Messung darstellt.

Tab. 7.17: Additive Indizes: Vertrauen in Implementationsinstitutionen und Eingebundenheit in informelle/formelle Netzwerke; Gruppe 1: Objektiv/Standard (N=50)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Vertrauen in Implementationsinstitutionen	3,58	,75835	-1,160	1,944	1	5
Anzahl der Mitgliedschaften in formellen Netzwerken	1,50	1,01519	,549	-,019	0	4
Eingebundenheit in informelle Netzwerke	3,44	,73290	,054	-,171	2	5

^a Arithmetisches Mittel

^b Standardabweichung

^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)

^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Betrachtet man die Durchschnittswerte der verschiedenen Variablen über die einzelnen Experimentalgruppen hinweg, zeigt sich, dass es wechselseitig keine Ausreißer gibt. Die Probanden scheinen hinsichtlich der erfassten Merkmale über die Gruppengrenzen hinweg ein vergleichsweise homogenes Kollektiv darzustellen. Das Vertrauen in Implementationsinstitutionen liegt dabei im Durchschnitt oberhalb der Skalenmitte, sodass von einem tendenziell stark ausgeprägten Vertrauen dieser Art ausgegangen werden kann. Ähnliches zeigt sich hinsichtlich der Eingebundenheit in informelle Netzwerke. Außerdem sind die Probanden im Gruppenmittel Mitglied in 1,46 formellen Netzwerken. Dies erscheint allerdings als ein relativ niedriger Wert. Am häufigsten nannten die insgesamt 204 Probanden dabei die Mitgliedschaft in einem Sport- oder Freizeitverein (48%), darauf folgte die Mitgliedschaft in einem kulturellen Verein (23,5%) und die Mitgliedschaft in einer Partei (15,2%). In keinem formellen Netzwerk Mitglied zu sein, gaben 14,7% der Befragten an.

Tab. 7.18: Additive Indizes: Vertrauen in Implementationsinstitutionen und Eingebundenheit in informelle/formelle Netzwerke; Gruppe 2: Standard/Objektiv (N=50)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Vertrauen in Implementationsinstitutionen	3,42	,75835	-,303	-,399	2	5
Anzahl der Mitgliedschaften in formellen Netzwerken	1,64	1,15635	,510	-,230	0	4
Eingebundenheit in informelle Netzwerke	3,16	,79179	-,555	1,006	1	5

^a Arithmetisches Mittel

^b Standardabweichung

^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)

^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Der Eindruck einer insgesamt homogenen Stichprobe verstärkt sich bei Betrachtung der Standardabweichungen. Vergleicht man die Kennwerte der einzelnen Variablen über die verschiedenen Experimentalgruppen, ergeben sich jeweils ähnliche Werte. Mit anderen Worten: Die Angaben hinsichtlich des Vertrauens in Implementationsinstitutionen und hinsichtlich der Eingebundenheit in informelle bzw. formelle Netzwerke der Probanden variieren in den unterschiedlichen Gruppen in annähernd gleicher Art und Weise.

Tab. 7.19: Additive Indizes: Vertrauen in Implementationsinstitutionen und Eingebundenheit in informelle/formelle Netzwerke; Gruppe 3: Objektiv (N=52)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Vertrauen in Implementationsinstitutionen	3,3269	,85683	-,309	,043	1	5
Anzahl der Mitgliedschaften in formellen Netzwerken	1,2308	,96234	,747	,328	0	4
Eingebundenheit in informelle Netzwerke	3,2500	,73764	-,133	-,576	2	5

^a Arithmetisches Mittel^b Standardabweichung^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Die Verteilung des Vertrauens in Implementationsinstitutionen fällt in allen Versuchsgruppen rechtssteil aus, was nochmals unterstreicht, dass die Probanden über ein tendenziell ausgeprägtes Vertrauen gegenüber dieser Klasse von Institutionen verfügen.

Die Verteilung der Mitgliedschaften in formellen Netzwerken fällt dagegen grundsätzlich linkssteil aus. Dies stellt ebenfalls eine Bestätigung der Ergebnisse der Mittelwertanalyse dar.

Die Eingebundenheit in informelle Netzwerke weist dagegen je nach Versuchsgruppe eine rechtssteile Tendenz oder annähernde Symmetrie auf. Dies ist insofern überraschend, als dass sowohl die Mittelwerte als auch die zugehörigen Streuungen eine vergleichsweise geringe Varianz zwischen den Gruppen aufweisen. Das Ergebnis lässt allerdings vermuten, dass die Probanden prinzipiell eher gut in informelle Netzwerke eingebunden sind.

Tab. 7.20: Additive Indizes: Vertrauen in Implementationsinstitutionen und Eingebundenheit in informelle/formelle Netzwerke; Gruppe 4: Standard (N=52)

	\bar{x}_i^a	σ_i^b	Sch_i^c	Ex_i^d	Min	Max
Vertrauen in Implementationsinstitutionen	3,4038	,86907	-,531	,045	1	5
Anzahl der Mitgliedschaften in formellen Netzwerken	1,4902	1,12022	,781	,051	0	4
Eingebundenheit in informelle Netzwerke	3,1731	,70631	,088	-,231	2	5

^a Arithmetisches Mittel^b Standardabweichung^c Schiefe ($Sch_i < 0 \rightarrow$ rechtssteile Verteilung; $Sch_i > 0 \rightarrow$ linkssteile Verteilung) (Bühner (2006), S. 88)^d Exzess bzw. Kurtosis ($Ex_i < 0 \rightarrow$ breitgipflige Verteilung; $Ex_i > 0 \rightarrow$ schmalgipflige Verteilung) (Bühner (2006), S. 88f)

(Quelle: Eigene Berechnung)

Hinsichtlich der Verteilungsform lassen sich keine gruppenübergreifenden Systematiken entdecken.

7.2.3 Validitätsanalyse

Wie bereits im Rahmen der Pilotstudie beschrieben⁷⁷, gibt das Gütekriterium „Validität“ an, ob mit einem Test tatsächlich die anvisierte Zieldimension abgebildet wird. Im nachfolgenden Abschnitt soll nun eine erneute umfassende Validitätsprüfung der in der Pilotstudie entwickelten Objektiv-Skala bzw. Kurzskala zur Erfassung Generalisierten Vertrauens anhand der durch das Surveyexperiment der Hauptstudie gewonnenen Daten vollzogen werden. Als

⁷⁷ Vgl. Abschnitt „6.2.8 Validitätsanalyse“

7. Hauptstudie

Bewertungsgrundlage werden dabei die von BÜHNER vorgeschlagenen Grenzwerte für die verschiedenen Validitätsformen herangezogen. (Bühner (2011), S. 61ff, S. 80f)

7.2.3.1 Inhaltsvalidität

Die Inhaltsvalidität der Objektiv-Skala ergibt sich aus der theoriegeleiteten Operationalisierung des Konstruktes „Generalisiertes Vertrauen“ und den faktorenanalytischen Untersuchungen im Rahmen der Pilotstudie.

7.2.3.2 Augenscheinvalidität (Face-Validity)

Objektive Tests sind u.a. durch eine Verschleierung der Messintention gekennzeichnet. Die Maskierung der Zieldimension wird dabei durch eine geringe Augenscheinvalidität der verwendeten Items und eine Coverstory gewährleistet. Um zu prüfen, ob für die innerhalb der Hauptstudie verwendeten Objektiv-Skalen die angestrebte geringe Augenscheinvalidität gegeben war, wurden die Probanden im Anschluss an die Bearbeitung der jeweiligen Objektiv-Skala gebeten, die folgende Frage zu beantworten:

„Was glauben Sie, wurde mit Hilfe der vorangegangenen Statements abgefragt bzw. welches Konstrukt sollte ihrer Meinung mit den vorangegangenen Aussagen gemessen werden?“

Eine erfolgreiche Verschleierung der Messintention erscheint dann als geglückt, wenn die Mehrzahl der Befragten hierbei ein von Generalisiertem Vertrauen verschiedenes Konstrukt als Zieldimension angibt.

Tab. 7.21: Prüfung der Augenscheinvalidität: Gruppe 1: Objektiv/Standard (N = 50)

Ausprägung	Häufigkeit	Gültige %
Vertrauen	12	24,0
Anderes Konstrukt, davon:	38	76,0
Weltbild/Menschenbild	21	42,0
Sozialverhalten	9	18,0
Lebenszufriedenheit	2	4,0
Charakterdispositionen	3	6,0
Sonstige	3	6,0
Fehlend	---	---

(Quelle: Eigene Berechnung)

Tab. 7.22: Prüfung der Augenscheinvalidität: Gruppe 2: Standard/Objektiv (N = 50)

Ausprägung	Häufigkeit	Gültige %
Vertrauen	24	48,0
Anderes Konstrukt, davon:	26	52,0
Weltbild/Menschenbild	17	34,0
Sozialverhalten	3	6,0
Lebenszufriedenheit	2	4,0
Charakterdispositionen	2	4,0
Sonstige	2	4,0
Fehlend	---	---

(Quelle: Eigene Berechnung)

Tab. 7.23: Prüfung der Augenscheinvalidität: Gruppe 3: Objektiv (N = 52)

Ausprägung	Häufigkeit	Gültige %
Vertrauen	11	22,0
Anderes Konstrukt, davon:	39	78,0
Weltbild/Menschenbild	20	40,0
Sozialverhalten	13	26,0
Lebenszufriedenheit	4	8,0
Charakterdispositionen	0	---
Sonstige	2	4,0
Fehlend	2	---

(Quelle: Eigene Berechnung)

Über alle Versuchsgruppen hinweg, in denen Generalisiertes Vertrauen per Objektiv-Skala abgefragt wurde, entfiel bei der Frage nach der Zieldimension der Objektiv-Skala die Mehrzahl der gültigen Antworten auf Konstrukte, die von Vertrauen verschieden waren. Während in den *Gruppen 1* und *3* jeweils ein deutlicher Unterschied zwischen den Anteilen der Antwortkategorien „Vertrauen“ und „Anderes Konstrukt“ besteht, ist diese Differenz in *Gruppe 2* allerdings weit weniger stark ausgeprägt. Dieser Umstand ist mit großer Wahrscheinlichkeit auf die Bearbeitungsreihenfolge von Objektiv- und Standard-Skala innerhalb dieser Gruppe zurückzuführen. So wurde den Probanden in *Gruppe 2* zunächst die Standard-Skala und erst im Anschluss die Objektiv-Skala präsentiert. Da im Rahmen der Standard-Skala die Zieldimension „Generalisiertes Vertrauen“ offengelegt wurde, könnte ein Priming-Effekt für die gesteigerte Identifikation der Messintention der Objektiv-Skala verantwortlich sein. Wie in nachfolgenden Analyseschritten gezeigt werden kann, hatte dieser Umstand aber keinen signifikanten Einfluss auf das Antwortverhalten der Probanden.

Als Zwischenergebnis lässt sich konstatieren, dass in der Tendenz für die innerhalb der Objektiv-Skala verwendeten Vertrauensitems in allen Versuchsgruppen eine geringe Augenscheinvalidität bezüglich der Zieldimension gewährleistet werden konnte. So liegt der Anteil der Probanden, die die Messintention nicht erkannt hatten, in allen Gruppen oberhalb des Anteils derer, die das Konstrukt „Generalisiertes Vertrauen“ identifizieren konnten.

Das Ergebnis der Augenscheinprüfung spiegelt sich im Wesentlichen auch in den jeweiligen Konfidenzintervallen der Stichprobenkennwerte wider. So lässt sich für die *Versuchsgruppen 1* und *3* keine Überlappung der Streubereiche der wahren Anteilswerte beobachten. Die obere Intervallgrenze für den Anteilswert der Kategorie „Vertrauen“ liegt jeweils weit unterhalb der unteren Intervallgrenze für den Anteilswert der Kategorie „Anderes Konstrukt“. Hieraus lässt sich schließen, dass sich die wahren Anteilswerte voneinander unterscheiden. Außerdem lässt sich an den Intervallgrenzen ablesen, dass der wahre Anteil der Kategorie „Vertrauen“ weit unterhalb des wahren Anteils der Kategorie „Anderes Konstrukt“ liegt. Mit anderen Worten: Der wahre Anteil der Kategorie „Vertrauen“ variiert in einem niedrigeren Wertebereich als der wahre Anteil der Kategorie „Anderes Konstrukt“. Somit kann angenommen werden, dass mehr Probanden nicht in der Lage sind, die Messintention zu erkennen, als Probanden in der Lage sind, sie zu identifizieren. Hieraus lässt sich schließen, dass in diesen Gruppen die Verschleierung der Messintention der Objektiv-Skala tendenziell erfolgreich war.

Tab. 7.24: Konfidenzintervalle der Stichprobenkennwerte

Gruppe 1: Objektiv/Standard ($\sigma_{\bar{x}} = 0,06101$)^a	
„Vertrauen“	Konf _{99%} = $0,0765 \leq p \leq 0,4035$
„Anderes Konstrukt“	Konf _{99%} = $0,5965 \leq p \leq 0,9235$
Gruppe 2: Standard/Objektiv ($\sigma_{\bar{x}} = 0,07137$)^a	
„Vertrauen“	Konf _{99%} = $0,2887 \leq p \leq 0,6713$
„Anderes Konstrukt“	Konf _{99%} = $0,3287 \leq p \leq 0,7113$
Gruppe 3: Objektiv ($\sigma_{\bar{x}} = 0,05918$)^a	
„Vertrauen“	Konf _{99%} = $0,0614 \leq p \leq 0,3786$
„Anderes Konstrukt“	Konf _{99%} = $0,6214 \leq p \leq 0,9386$

^a Standardstichprobenfehler (Bortz (2005), S. 104)

(Quelle: Eigene Berechnung)

Ein anderes Resultat ergibt sich für die *zweite Versuchsgruppe*. Hier bewegen sich die wahren Anteilswerte der Antwortkategorie „Vertrauen“ in einem Bereich von 28,87% bis 67,13% und die Anteile der Antwortkategorie „Anderes Konstrukt“ zwischen 32,87% und 71,13%. Es findet somit eine Überschneidung der Konfidenzintervalle statt. Daher ist anzunehmen, dass sich die wahren Anteilswerte der Antwortkategorien „Vertrauen“ und „Anderes Konstrukt“ nicht unterscheiden oder sogar eine im Vergleich zur in *Versuchsgruppe 2* gefundenen Verteilung umgekehrte Relation der Anteilswerte der Antwortkategorien bestehen könnte. Entsprechend scheint in dieser Gruppe für die Objektiv-Skala keine Verschleierung der Messintention garantiert zu sein.

Sieht man von *Versuchsgruppe 2* ab, so kann für die Items der Objektiv-Skala durchaus eine geringe Augenscheinvalidität bezüglich der Zieldimension „Generalisiertes Vertrauen“ als gegeben angesehen werden. So haben in *Versuchsgruppe 1* und *2* jeweils nur knapp ein Viertel der Befragten die Messintention erkannt, während ca. Dreiviertel der Befragten die Messintention nicht feststellen konnten. Die Relation von Identifikation und Nicht-Identifikation der Messintention deckt sich dabei auch mit dem Ergebnis der Pilotstudie. Hier hatten 20,5% von N=220 die Messintention identifiziert, während 79,5% die Messintention nicht erkannt hatten. Vor diesem Hintergrund sind die Abweichungen in *Gruppe 2* (ca. 50:50-Relation) als Ausreißer zu bewerten, der mit der Bearbeitungsreihenfolge der verwendeten Vertrauensskalen zu begründen ist. Es ist also anzunehmen, dass eine alleinige Präsentation der Objektiv-Skala - ohne die Offenlegung potentieller Hinweise auf deren Zieldimension - in den meisten Fällen eine erfolgreiche Verschleierung der Messintention ermöglicht.

Im Anschluss an die Frage nach der Messintention der Objektiv-Skala wurden die Probanden außerdem gefragt, welches der Items den Hinweis auf die subjektiv angenommene Zieldimension geliefert hat:

„Was hat Sie zu Ihrer Antwort auf die vorangegangene Frage bewogen bzw. woran glauben Sie die Intention des Fragebogens erkannt zu haben?“

Nachfolgend werden nur die Nennungen der Fälle betrachtet, die „Generalisiertes Vertrauen“ als Zieldimension angegeben hatten. Dabei ergaben sich je nach Versuchsgruppe unterschiedliche Schwerpunktsetzungen hinsichtlich der genannten Items und Unterschiede in den absoluten Häufigkeiten von Item-Nennungen. So wurden in *Versuchsgruppe 2* insgesamt deutlich mehr Item-Nennungen gezählt als in den anderen Gruppen. Dieser Effekt ist wohl ein

7. Hauptstudie

weiteres Mal auf das Priming infolge der Skalenabfolge innerhalb von *Versuchsgruppe 2* zurückzuführen.

Tab. 7.25: Item als Identifikationshinweis für Zieldimension „Generalisiertes Vertrauen“

Item	Anzahl der Item-Nennungen			Gesamt
	Objektiv/Standard	Standard/Objektiv	Objektiv	
v2.1	0	5	1	6
v2.2	4	5	3	12
v2.3	1	7	3	11
v2.4	1	5	1	7
v2.5	2	4	3	9
v2.6	3	5	2	10
v2.7	3	9	0	12
v2.8	1	2	0	3
v2.9	2	2	2	6
v2.10	2	9	3	14
v2.11	1	4	1	6
v2.12	3	3	1	7
v2.13	4	8	2	14
v2.14	4	7	2	13
Gesamt	31	75	24	130

(Quelle: Eigene Berechnung)

Über alle Versuchsgruppen hinweg betrachtet wurden die Items v2.10, v2.13 und v2.14 am häufigsten als Identifikationshinweise für die Zieldimension „Generalisiertes Vertrauen“ genannt. Der Grund hierfür ist vermutlich in der Nähe deren Formulierungen zum Wortlaut der Items der Standard-Skala zu suchen. So scheint eine Ähnlichkeit von v2.10 mit der Hilfsbereitschaftskomponente, von v2.13 mit der Vertrauenskomponente und von v2.14 mit der Fairnesskomponente⁷⁸ der *Faith in People-Scale* zu bestehen. Dabei ist allerdings zu beachten, dass das Gros der Nennungen aus *Versuchsgruppe 2* stammt.

Ähnlich häufige Nennungen in allen Gruppen lassen sich dagegen für die Items v2.2, v2.5, v2.6, v2.9 und v2.12 feststellen. Dabei wurden insbesondere v2.2, v2.5 und v2.6 überdurchschnittlich oft genannt. Diese Items entsprechen, folgt man der in der Pilotstudie ermittelten Faktorstruktur, den Komponenten der *Faith in People-Scale*.

Tatsächlich lässt sich aus der Verteilung der Item-Nennungen über die Versuchsgruppen kein nachhaltiges Muster ableiten. Dafür sind die Item-Nennungen innerhalb der *Versuchsgruppen 1* und *2* zu differenziert verteilt bzw. keines der Items kann hier besonders hervorstechen. Die überdurchschnittliche Nennung bestimmter Items innerhalb von *Gruppe 2* ist mit aller Voraussicht einem *Priming*-Effekt anzulasten. Mit anderen Worten: Es kann kein spezielles Item isoliert werden, das im Besonderen zur Aufdeckung der Messintention beigetragen hat. Gleichzeitig lässt sich aber konstatieren, dass die Demaskierung der Zieldimension der Objektiv-Skala vornehmlich entlang von Items verläuft, die den verschiedenen Komponenten der *Trust in People-Scale* bzw. dem Standardverfahren zur Erfassung Generalisierten Vertrauens entsprechen. Dies sollte allerdings mit Vorsicht als potientiell Antwortmuster gewertet werden, da sich im Rahmen der Pilotstudie gezeigt hat, dass die Faktorstruktur der Objektiv-Skala bereits an die inhaltliche Struktur (Vertrauens-, Hilfsbereitschafts- und Fairnesskomponente) der *Trust in People-Scale* angelehnt ist.

⁷⁸ Entspricht nicht der Item-Faktor-Zuordnung infolge der Pilotstudie vgl. Abschnitt „6.2.3 Prüfung auf Eindimensionalität (Objektiv-Skala)“ bzw. „6.2.6 Prüfung auf Eindimensionalität (Objektiv-Kurzskala)“

7.2.3.3 Konstruktvalidität

Die Konstruktvalidität stellt das wichtigste Validitätskriterium der Testtheorie dar und lässt sich in die Teilkomponenten *konvergente*, *divergente* und *faktorielle Validität* differenzieren. (Bühner (2011), S. 63f; Moosbrugger/Kelava (2012), S. 153ff) Sie beschreibt im Allgemeinen, „wie angemessen ein Test das erfasst, was er zu messen beansprucht“ (Bühner (2011), S. 63). Von einer Bestimmung der divergenten Validität wurde sowohl in der Pilotstudie als auch der Hauptstudie abgesehen, da die Aufnahme von Tests anderer Gültigkeitsbereiche zu einer inhaltlichen Überladung der eingesetzten Fragebogen geführt hätte. Während die Pilotstudie lediglich eine rudimentäre Bestimmung der *faktoriellen Validität* der Objektiv-Skala zuließ, scheiterte eine Bestimmung der *konvergenten Validität* an der Tatsache, dass aufgrund der Länge des eingesetzten Fragebogens keine parallele Messung der Zieldimension „Generalisiertes Vertrauen“ mit Hilfe einer alternativen Vertrauensskala realisiert wurde. Eine erneute Bestimmung der *faktoriellen Validität* ist im Rahmen der Hauptstudie nicht möglich, da die Fallzahl je Versuchsgruppe für die hierzu notwendigen Analyseschritte nicht ausreichend groß ist. Eine Kombinierung der einzelnen Stichproben verbietet sich vor dem Hintergrund, dass es sich bei den verschiedenen Versuchsgruppen des Surveyexperimentes um jeweils voneinander unabhängige Stichproben handelt.

Die Bestimmung der *konvergenten Validität* soll nun anhand der *Versuchsgruppen 1* und *2* (PR_1/PT_1 (Objektiv/Standard) u. PR_2/PT_2 (Standard/Objektiv)) nachgeholt werden. Als Paralleltest zur Erfassung Generalisierten Vertrauens per Objektivem Test wurde den Probanden in diesen Gruppen die deutsche Version der *Trust-in-People-Skala*, die innerhalb der Hauptstudie als Standard-Skala bezeichnet wird, vorgelegt. Als Validitätskoeffizient wird der *Produkt-Moment-Korrelationskoeffizient von Pearson* verwendet. Von *konvergenter Validität* kann dann ausgegangen werden, wenn Tests gleicher oder ähnlicher Gültigkeitsbereiche einen hohen Zusammenhang aufweisen.

Tab. 7.26: Bivariate Korrelationen:

Summenscore Objektiv-Skala (PR_1) und Summenscore Standard-Skala (PT_1) (N=50)

	PT_1^a	PR_1^b	$PR_{1(KS)}^c$
PT_1^a	1	---	---
PR_1^b	,605**	1	---
$PR_{1(KS)}^c$,675**	,798**	1

^a Standard-Skala

^b Objektiv-Skala

^c Objektiv-Kurzskala

* $p < .05$ (zweiseitig), ** $p < .01$ (zweiseitig)

(Quelle: Eigene Berechnung)

In *Versuchsgruppe 1 Objektiv/Standard* ergeben sich sowohl zwischen der Objektiv-Skala und der Standard-Skala als auch zwischen der Objektiv-Kurzskala und der Standard-Skala stark positive Zusammenhänge. In beiden Fällen liegt die Korrelation oberhalb des für konvergente Validität angenommenen Schwellenwertes von $r > .50$. Es darf deshalb für beide Objektiv-Skalen angenommen werden, dass konvergente Validität gegeben ist bzw. die Standard-Skala und die Objektiv-Skalen die gleiche Zieldimension abbilden.

Tab. 7.27: Bivariate Korrelationen:
Summenscore Objektiv-Skala (PT₂) und Summenscore Standard-Skala (PR₂) (N=50)

	PR ₂ ^a	PT ₂ ^b	PT _{2(KS)} ^c
PR ₂ ^a	1	---	---
PT ₂ ^b	,502**	1	---
PT _{2(KS)} ^c	,603**	,640**	1

^a Standard-Skala

^b Objektiv-Skala

^c Objektiv-Kurzskala

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: Eigene Berechnung)

Auch für *Versuchsgruppe 2 Standard/Objektiv* ergeben sich ähnlich starke positive Zusammenhänge zwischen den Objektiv-Skalen und der Standardskala, die ebenfalls oberhalb des Schwellenwertes für konvergente Validität liegen. Somit scheint konvergente Validität vorzuliegen und es darf davon ausgegangen werden, dass beide Skalentypen das gleiche theoretische Konstrukt erfassen.

Bezieht man das Radiusproblem in die Analyse der Konstruktvalidität ein, wonach nur dann tatsächlich Generalisiertes Vertrauen gemessen wird, wenn generalisierte Andere bzw. die *outgroup* als Vertrauensreferenz genannt werden, so sinken die Korrelationswerte zwar nur leicht ab⁷⁹, liegen aber teilweise unterhalb des Schwellenwertes für die Konstruktvalidität. Unabhängig davon handelt es sich jeweils um signifikante Zusammenhänge. Hierbei ist allerdings zu beachten, dass sich unter diesen Bedingungen die in die Analyse eingehende Fallzahl im Vergleich zu der oben berichteten Analyse in *Versuchsgruppe 1* mehr als halbiert und in *Versuchsgruppe 2* nahezu halbiert.

Die leichten Unterschiede hinsichtlich der Skalen-Korrelationen zwischen den beiden Versuchsgruppen sind möglicherweise auf die jeweils gespiegelte Präsentationsreihenfolge der Skalen und daraus resultierender Unterschiede in der Interpretation der *demand characteristics* zurückzuführen. So besteht die Möglichkeit, dass die Probanden aufgrund der Offenlegung der Zieldimension „Generalisiertes Vertrauen“ durch die Präsentation der Standard-Skala in PR₂ bei der Bearbeitung der Objektiv-Skala in PT₂ davon ausgingen, dass nicht zweimal das gleiche Konstrukt abgefragt würde bzw. es sich bei der zweiten Messung um ein fundamental verschiedenes Konstrukt handeln müsste und sich deshalb in *Versuchsgruppe 2* ein etwas geringerer Zusammenhang zwischen den Skalentypen ergibt.

Geht man von der Prämisse aus, dass es sich bei der Erfassung von Generalisiertem Vertrauen mit Hilfe der Standard-Skala um eine valide Messung handelt, lässt sich aus den vorliegenden Ergebnissen nun schließen, dass auch die Objektiv-Skalen eine valide Messung Generalisierten Vertrauens ermöglichen. Ferner lässt sich aus der vergleichsweise höheren Korrelation von Objektiv-Kurzskala und Standard-Skala ableiten, dass die Objektiv-Kurzskala eine bessere Passung mit der Standard-Skala aufweist als die Objektiv-Skala und deshalb der Objektiv-Skala bei der Abbildung Generalisierten Vertrauens überlegen ist. Diese Ähnlichkeit ist vermutlich der Inhaltkongruenz der Faktorstrukturen von Objektiv-Kurzskala und Standard-Skala geschuldet.

⁷⁹ Vgl. Anhang I: Validität

7.2.3.4 Kriteriumsvalidität

Kriteriumsvalidität liegt vor, wenn ein Test mit einem oder mehreren Kriterien korreliert, zu denen die Zieldimension naturgemäß in Beziehung stehen muss. Die Kriteriumsvalidität kann dabei in Vorhersagevalidität, Übereinstimmungsvalidität, retrospektive Validität und inkrementelle Validität unterschieden werden. (Bühner (2011), S. 63; Moosbrugger/Kelava (2012), S. 164ff)

Innerhalb der Hauptstudie wird ausschließlich die Übereinstimmungsvalidität geprüft. Als externe Kriterien werden hierbei das subjektive Wohlbefinden, das Vertrauen in Implementationsinstitutionen, die Eingebundenheit in informelle Netzwerke und die Mitgliedschaft in formellen Netzwerken herangezogen. Die genannten Kriterien wurden jeweils parallel zur Vertrauensmessung erfasst. Es wird dabei davon ausgegangen, dass jedes der externen Kriterien positiv mit Generalisiertem Vertrauen korreliert ist. So bewirkt subjektives Wohlbefinden in der Regel eine optimistischere Weltwahrnehmung und damit auch ein positiveres Bild von den übrigen Gesellschaftsmitgliedern. In der Folge erhöht sich die Bereitschaft, sowohl bekannten als auch fremden Personen gegenüber vertrauensvoller eingestellt zu sein. Mit dem Anstieg des subjektiven Wohlbefindens ergibt sich somit ein höheres Generalisiertes Vertrauen.

Implementationsinstitutionen garantieren bzw. erschaffen durch die Überwachung und Gewährleistung einer Rechtsordnung bzw. institutioneller Regeln eine soziale Ordnung, die die Gesellschaftsmitglieder verpflichtet, innerhalb einer spezifischen normativen Rahmung zu handeln. Vor dem Hintergrund dieser Rahmung kann das Handeln anderer Personen besser antizipiert werden, was die Vergabe von Vertrauen erleichtert.

Positive Kooperationserfahrungen innerhalb informeller oder formeller sozialer Netzwerke bzw. die Eingebundenheit in Freundschaften und soziale Vereinigungen bewirken in der Regel eine Steigerung der Vertrauensbereitschaft. Negative Erlebnisse können dagegen zum Verfall der Vertrauensbereitschaft beitragen. Insbesondere im Rahmen formeller Netzwerke werden dabei auch oft Vertrauensbeziehungen mit fremden Personen eingegangen. Die gesammelten Erfahrungen werden in der Folge auf den gesamtgesellschaftlichen Kontext generalisiert und manifestieren sich in einer individuellen Ausstattung mit Generalisiertem Vertrauen bzw. Misstrauen.

Da es sich bei den verschiedenen Versuchsgruppen um voneinander unabhängige Stichproben handelt, wird auch im Zuge der Bestimmung der konkurrenten Validität jede Versuchsgruppe einzeln betrachtet. Als Grenzwert für die Kriteriumsvalidität schlägt BÜHNER $r_{tc} > .20$ vor. (Bühner (2011), S. 80)

7.2.3.4.1 Gruppe 1: Objektiv/Standard

Innerhalb von *Gruppe 1: Objektiv/Standard* ergeben sich je nach betrachtetem Kriterium und betrachteter Skala erhebliche Diskrepanzen hinsichtlich der Kriteriumsvalidität. So kann für die Objektiv-Skala und die Objektiv-Kurzskala über die Kriterien „Vertrauen in Implementationsinstitutionen“ und „Eingebundenheit in informelle Netzwerke“ jeweils Kriteriumsvalidität festgestellt werden. Über die übrigen Kriterien lässt sich hinsichtlich der Objektiv-Skala allerdings gar keine und hinsichtlich der Objektiv-Kurzskala kaum Kriteriumsvalidität feststellen. Für beide Skalen liegen die Korrelationen unterhalb der Schwelle von $r_{tc} > .20$. Für die Standard-Skala lässt sich indes anhand aller Kriterien

Übereinstimmungsvalidität feststellen. Einzige Ausnahme bildet hierbei das Vertrauen in Implementationsinstitutionen. Über alle Skalen hinweg betrachtet scheint die Eingebundenheit in informelle Netzwerke den größten Einfluss auf Generalisiertes Vertrauen auszuüben. Dies lässt darauf schließen, dass für die Entwicklung einer Generalisierten Vertrauensdisposition ein stabiles soziales Umfeld von erheblicher Bedeutung ist.

Die geringen Korrelationen zwischen den Objektiv-Varianten zur Vertrauensmessung und der aktuellen Zufriedenheit bzw. der allgemeinen Lebenszufriedenheit sind dabei wohl vornehmlich darauf zurückzuführen, dass die Probanden auf den Vertrauensskalen eine Tendenz zur Mitte aufweisen. Zwar gibt die Mehrzahl der Probanden eine hohe aktuelle Zufriedenheit und eine hohe allgemeine Lebenszufriedenheit an, gleichzeitig geht diese Angabe aber in den meisten Fällen mit einem moderaten Generalisierten Vertrauen zusammen. Bei der Objektiv-Skala kommt außerdem hinzu, dass nur sehr wenige Probanden ein Vertrauensniveau oberhalb der Mittelkategorie erreichen. Im Vergleich dazu ist das Antwortverhalten auf der Standard-Skala differenzierter bzw. es wird im Allgemeinen ein höheres Vertrauenslevel angegeben. Entsprechend sind hier stärkere positive Korrelationen möglich.

Der Unterschied zwischen Objektiv- und Standard-Skala hinsichtlich des Zusammenhangs mit der Anzahl der Mitgliedschaften in formellen Netzwerken lässt sich darauf zurückführen, dass sich für die Verteilung der Probanden mit wenigen oder vielen Netzwerkmitgliedschaften auf die Antwortkategorien der Objektiv-Skala kein klares Muster ergibt. Dagegen ist für die Standard-Skala klar ersichtlich, dass wenige Netzwerkmitgliedschaften mit einem niedrigeren generalisierten Vertrauen einhergehen, was den theoretischen Vorüberlegungen entspricht.

Über alle Kriterien hinweg weisen die Objektiv-Kurzskala und die Standard-Skala vergleichsweise höhere Korrelationen als die Objektiv-Skala auf. Dies stellt einen Hinweis darauf dar, dass die Objektiv-Kurzskala eine validere Abbildung Generalisierten Vertrauens liefert als die Objektiv-Skala. Gleichzeitig kann aber festgestellt werden, dass die Kriteriumsvalidität der Standard-Skala noch stärker ausgeprägt ist. Sie scheint im Vergleich die beste Messvariante zu sein.

7.2.3.4.2 Gruppe 2: Standard/Objektiv

Innerhalb der zweiten Versuchsgruppe lässt sich lediglich anhand der Eingebundenheit in informelle Netzwerke Kriteriumsvalidität für die Objektiv-Skala konstatieren. Die übrigen Kriterien weisen zu niedrige oder sogar negative Korrelationen auf. Die Objektiv-Kurzskala und die Standard-Skala können dagegen sowohl durch die Eingebundenheit in informelle Netzwerke als auch durch das Vertrauen in Implementationsinstitutionen als validiert angesehen werden. Für die übrigen Kriterien ergeben sich auch hier zu niedrige oder negative Korrelationen.

Wie bereits in *Gruppe 1* scheint auch für die Probanden in *Gruppe 2* die Eingebundenheit in informelle Netzwerke der beste Prädiktor für Generalisiertes Vertrauen zu sein. Die Mitgliedschaft in formellen Netzwerken scheint dagegen zumindest in dieser Gruppe einen den theoretischen Annahmen diametral entgegengesetzten Effekt zu haben. Personen, die sich weniger in formellen Netzwerken engagieren, weisen tendenziell ein niedrigeres Generalisiertes Vertrauen auf. Ähnliches gilt für das subjektive Wohlbefinden. Dieses hat entweder keinen oder einen negativen Effekt. Während die allgemeine Lebenszufriedenheit kaum bis gar keine Rolle spielt, wirkt sich eine unglückliche Tagesverfassung positiv auf den Vertrauenslevel aus.

In *Gruppe 2* sind die Objektiv-Kurzskala und die Standard-Skala in etwa gleich auf, was die Übereinstimmungsvalidität angeht. Beide korrelieren stärker mit den externen Kriterien als die

Objektiv-Skala. Sie scheinen dementsprechend auch eine validere Erfassung Generalisierten Vertrauens zu ermöglichen.

7.2.3.4.3 Gruppe 3: Objektiv

Da im Rahmen der *dritten Versuchsgruppe* keine Abfrage Generalisierten Vertrauens per Standard-Skala stattgefunden hat, kann im folgenden Abschnitt ausschließlich die Übereinstimmungsvalidität für die Objektiv-Varianten bestimmt werden.

Dabei lässt sich für die Objektiv-Skala ausschließlich anhand der Eingebundenheit in informelle Netzwerke Validität ableiten. Alle übrigen Kriterien weisen zu niedrige Korrelationen auf. Die Objektiv-Kurzskala weist hingegen zusätzlich eine ausreichend hohe Korrelation mit der aktuellen Zufriedenheit der Probanden auf. Insgesamt korreliert die Objektiv-Kurzskala allerdings deutlich höher mit den übrigen Kriterien als die Objektiv-Skala. Dies deutet daraufhin, dass die Objektiv-Kurzskala eine bessere Abbildung Generalisierten Vertrauens liefert als die Objektiv-Skala bzw. die Erfassung Generalisierten Vertrauens mit Hilfe der Objektiv-Kurzskala weist eine höhere Validität auf.

Dagegen scheint die Mitgliedschaft in formellen Netzwerken keine Rolle für das Vertrauenslevel der Probanden zu spielen. Vielmehr scheint wie bereits in *Versuchsgruppe 2* ein den theoretischen Annahmen leicht gegenläufiger Trend erkennbar zu sein. So gehen moderate und hohe Vertrauenswerte eher mit einer geringen Anzahl an Mitgliedschaften einher.

7.2.3.4.4 Gruppe 4: Standard

Aus *Gruppe 4* lassen sich ausschließlich Informationen über die konkurrente Validität der Standard-Skala gewinnen. Dabei kann über alle Kriterien hinweg Validität festgestellt werden. Den einzigen Ausreißer stellt hierbei die Anzahl der Mitgliedschaften in formellen Netzwerken dar. Allerdings liegt die Korrelation mit der Standard-Skala nur knapp unterhalb der Schwelle für die Kriteriumsvalidität von $r_{tc} > .20$. Aus diesem Ergebnis lassen sich zwei Erkenntnisse ziehen: (1) Die zur Bestimmung der Übereinstimmungsvalidität gewählten externen Kriterien besitzen Relevanz für die individuelle Ausstattung mit Generalisiertem Vertrauen, was einer Bestätigung der theoretischen Vorüberlegungen gleichkommt und (2) die Standard-Skala ermöglicht eine valide Abbildung der individuellen Ausstattung mit Generalisiertem Vertrauen.

7. Hauptstudie

Tab. 7.28: Übereinstimmungsvalidität/Konkurrente Validität in Gruppe 1: Objektiv/Standard (N = 50)

	Vertrauen in Implementationsinstitutionen	Anzahl der Mitgliedschaften in formellen Netzwerken	Eingebundenheit in informelle Netzwerke	Zufriedenheit/Glück aktuell	Lebenszufriedenheit allgemein	Generalisiertes Vertrauen (Objektiv-Skala)	Generalisiertes Vertrauen (Objektiv-Kurzskala)	Generalisiertes Vertrauen (Standard-Skala)
Vertrauen in Implementations-Institutionen	1	---	---	---	---	---	---	---
Anzahl der Mitgliedschaften in formellen Netzwerken	,093	1	---	---	---	---	---	---
Eingebundenheit in informelle Netzwerke	,339*	,274	1	---	---	---	---	---
Zufriedenheit/Glück aktuell	,211	,257	,189	1	---	---	---	---
Lebenszufriedenheit allgemein	,024	,233	,249	,488**	1	---	---	---
Generalisiertes Vertrauen (Objektiv-Skala)	,387**	,000	,409**	,006	,031	1	---	---
Generalisiertes Vertrauen (Objektiv-Kurzskala)	,295*	,126	,392**	,156	,195	,798**	1	---
Generalisiertes Vertrauen (Standard-Skala)	,173	,212	,509**	,319*	,225	,605**	,675**	1

*p<.05 (zweiseitig), **p<.01 (zweiseitig)
(Quelle: Eigene Berechnung)

Tab. 7.29: Übereinstimmungsvalidität/Konkurrente Validität in Gruppe 2: Standard/Objektiv (N = 50)

	Vertrauen in Implementationsinstitutionen	Anzahl der Mitgliedschaften in formellen Netzwerken	Eingebundenheit in informelle Netzwerke	Zufriedenheit/Glück aktuell	Lebenszufriedenheit allgemein	Generalisiertes Vertrauen (Objektiv-Skala)	Generalisiertes Vertrauen (Objektiv-Kurzskala)	Generalisiertes Vertrauen (Standard-Skala)
Vertrauen in Implementations-Institutionen	1	---	---	---	---	---	---	---
Anzahl der Mitgliedschaften in formellen Netzwerken	,199	1	---	---	---	---	---	---
Eingebundenheit in informelle Netzwerke	,022	-,092	1	---	---	---	---	---
Zufriedenheit/Glück aktuell	,061	-,145	,200	1	---	---	---	---
Lebenszufriedenheit allgemein	,178	,061	,319*	,102	1	---	---	---
Generalisiertes Vertrauen (Objektiv-Skala)	,155	-,062	,278	-,161	,004	1	---	---
Generalisiertes Vertrauen (Objektiv-Kurzskala)	,255	-,181	,235	-,014	,133	,640**	1	---
Generalisiertes Vertrauen (Standard-Skala)	,265	-,019	,332*	,060	,043	,502**	,603**	1

*p<.05 (zweiseitig), **p<.01 (zweiseitig)
(Quelle: Eigene Berechnung)

7. Hauptstudie

Tab. 7.30: Übereinstimmungsvalidität/Konkurrente Validität in Gruppe 3: Objektiv (N = 52)

	Vertrauen in Implementationsinstitutionen	Anzahl der Mitgliedschaften in formellen Netzwerken	Eingebundenheit in informelle Netzwerke	Zufriedenheit/Glück aktuell	Lebenszufriedenheit allgemein	Generalisiertes Vertrauen (Objektiv-Skala)	Generalisiertes Vertrauen (Objektiv-Kurzskala)
Vertrauen in Implementationsinstitutionen	1	---	---	---	---	---	---
Anzahl der Mitgliedschaften in formellen Netzwerken	,026	1	---	---	---	---	---
Eingebundenheit in informelle Netzwerke	,178	,193	1	---	---	---	---
Zufriedenheit/Glück aktuell	-,050	,159	,220	1	---	---	---
Lebenszufriedenheit allgemein	,164	,321*	,218	,493**	1	---	---
Generalisiertes Vertrauen (Objektiv-Skala)	,175	,059	,207	,132	,123	1	---
Generalisiertes Vertrauen (Objektiv-Kurzskala)	,182	,003	,226	,222	,175	,760**	1

*p<.05 (zweiseitig), **p<.01 (zweiseitig)
(Quelle: Eigene Berechnung)

Tab. 7.31: Übereinstimmungsvalidität/Konkurrente Validität in Gruppe 4: Standard (N = 51)

	Vertrauen in Implementationsinstitutionen	Anzahl der Mitgliedschaften in formellen Netzwerken	Eingebundenheit in informelle Netzwerke	Zufriedenheit/Glück aktuell	Lebenszufriedenheit allgemein	Generalisiertes Vertrauen (Standard-Skala)
Vertrauen in Implementationsinstitutionen	1	---	---	---	---	---
Anzahl der Mitgliedschaften in formellen Netzwerken	,372**	1	---	---	---	---
Eingebundenheit in informelle Netzwerke	,465**	,389**	1	---	---	---
Zufriedenheit/Glück aktuell	,213	,035	,055	1	---	---
Lebenszufriedenheit allgemein	,277*	,145	,281*	,446**	1	---
Generalisiertes Vertrauen (Standard-Skala)	,585**	,197	,351*	,314*	,416**	1

*p<.05 (zweiseitig), **p<.01 (zweiseitig)
(Quelle: Eigene Berechnung)

7.2.3.4.5 Zusammenfassung

Zunächst kann festgestellt werden, dass für alle untersuchten Skalen über alle Versuchsgruppen hinweg mindestens durch einen Prädiktor Kriteriumsvalidität gegeben ist. Jede der betrachteten Skalen stellt somit eine mehr oder weniger valide Möglichkeit zur Abbildung der individuellen Ausstattung mit Generalisiertem Vertrauen dar. Hierbei ist allerdings einzuschränken, dass nicht alle verwendeten Kriterien gleich gut zur Bestimmung der konkurrenten Validität beitragen und der jeweilige Einfluss von Versuchsgruppe zu Versuchsgruppe unterschiedlich stark ausgeprägt ist. Es kann aber nicht zuletzt aufgrund der Resultate aus *Versuchsgruppe 4* geschlossen werden, dass jedes der ausgewählten Kriterien ein Korrelat Generalisierten Vertrauens darstellt. Die theoretischen Vorannahmen hinsichtlich des Einflusses von subjektivem Wohlbefinden, sozialen Netzwerken und institutioneller Rahmung scheinen somit bestätigt.

Wie aus der Pilotstudie bekannt, entspricht die Faktorstruktur der Objektiv-Kurzskala im Wesentlichen der Faktorstruktur der Standard-Skala. Dies schlägt sich aller Voraussicht nach auch in der Übereinstimmungsvalidität der Objektiv-Kurzskala nieder. Ausgehend von den vorangegangenen Resultaten besitzt die Standard-Skala augenscheinlich eine im Vergleich zu den Objektiv-Skalen verbesserte Validität hinsichtlich des Konstruktes „Generalisiertes Vertrauen“. Die Kriteriumsvalidität der Objektiv-Kurzskala rangiert direkt dahinter. Dies ist vermutlich auf die Übereinstimmung der Faktorstruktur beider Skalen zurückzuführen. Die Objektiv-Skala bietet im Vergleich eine eher verminderte Abbildungsgüte.

7.2.4 Radiusproblem

Das Radiusproblem der Vertrauensforschung stellt eine der Kernproblematiken der Vertrauensforschung dar und beschäftigt sich mit der Frage, ob die gängigen zur Erfassung von Generalisiertem Vertrauen eingesetzten Instrumente klar unterscheiden können, auf welchen Personenkreis sich das gemessene Vertrauen bezieht. Das bedeutet, es wird untersucht, welche Art von Vertrauen (Partikulares od. Generalisiertes Vertrauen) mit Hilfe des Standardinstrumentes (*Trust in People-Scale*) abgebildet wird. Innerhalb der Standarditems wird der angepeilte Bezugsrahmen des zu messenden Vertrauens durch die Phrasen „*most people*“ und „*most of the time*“ definiert. Im Wesentlichen wird nun untersucht, ob die Phrasen „*most people*“ und „*most of the time*“ innerhalb der standardmäßig verwendeten Items von allen Befragten in gleicher Art und Weise verstanden werden. Sind die Referenzpersonen des Vertrauens tatsächlich generalisierte Andere oder Personen aus dem näheren Bekanntenkreis bzw. solche, mit denen man regelmäßig interagiert und über die man bereits Informationen besitzt?

Um zu prüfen, ob die Umsetzung der Erfassung von Vertrauen mittels „Objektiver Tests“ einen Beitrag zur Lösung des Radiusproblems leisten kann, wurde innerhalb der Hauptstudie in Anlehnung an STURGIS/SMITH und DELHEY/NEWTON/WELZEL jeweils im Anschluss an die Objektiv-Skala und im Anschluss an die Standard-Skala die folgende Frage gestellt (Delhey/Newton/Welzel (2011), S. 790ff; Sturgis/Smith (2010), S. 80; Welzel (2010), S. 162f):

„An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht?“

7. Hauptstudie

Durch die parallele Abfrage der Referenzgruppen für Standard- und Objektiv-Skala wird es möglich, einen direkten Vergleich des jeweils erfassten Vertrauensradius zu realisieren. Als Antwortalternativen standen den Probanden dabei Personengruppen zur Verfügung, die sowohl ihre *ingroup* als auch ihre *outgroup* repräsentierten. Als *ingroup* gilt dabei der Personenkreis zu dem die Befragten intime Beziehungen pflegen bzw. der ihnen persönlich bekannt ist und mit dem regelmäßig interagiert wird. Als *outgroup* gelten dagegen Personen, mit denen man erst einmal Kontakt hatte, die sozialen Gruppen angehören, in denen man selbst nicht Mitglied ist, oder Fremde.

Da mit Hilfe der *Trust in People-Scale* standardmäßig Generalisiertes Vertrauen erfasst werden sollte, äußert sich das Radiusproblem vornehmlich dadurch, dass eine nicht unerhebliche Anzahl der Befragten die *ingroup* als Referenz ihres Vertrauens angeben. Es kann dadurch angezweifelt werden, dass die Standard-Skala eine sortenreine Abbildung Generalisierten Vertrauens darstellt bzw. eine Vermischung von partikularem und Generalisiertem Vertrauen stattfindet. Objektive Tests würden dann zur Lösung des Radiusproblems beitragen, wenn sich der Anteil der Befragten, die die *ingroup* als Vertrauensreferenz angeben, signifikant vom Anteil der Befragten unterscheidet, die die *outgroup* als Vertrauensreferenz angeben. Im Idealfall wählt dabei die Mehrzahl der Probanden die *outgroup* als Bezugsrahmen. Gleichzeitig sollte der Anteil der Probanden, die die *outgroup* als Referenz wählen, für die Objektiv-Skala größer sein als für die Standard-Skala. Mit anderen Worten: Der Hybrid scheint dann einen Beitrag zur Lösung des Radiusproblems zu leisten, wenn sich der Anteil des Bezugsrahmens *ingroup* signifikant vom Anteil des Bezugsrahmens *outgroup* unterscheidet bzw. wenn alle Befragten nur einen der beiden Bezugsrahmen gewählt haben.

7.2.4.1 Gruppe 1: Objektiv/Standard

Nachfolgend werden die Angaben zum Vertrauensradius der Objektiv- und der Standard-Skala aus *Versuchsgruppe 1* betrachtet und verglichen.

Tab. 7.32: Vertrauensradius für Objektiv-Skala (PR₁) (N = 50)

Ausprägung	Häufigkeit	Gültige %	Ausprägung	Häufigkeit	Gültige %
Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben. ^a	9	18,4	<i>ingroup</i>	20	40,8
Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren. ^b	11	22,4			
Personen, die Sie erst ein einziges Mal getroffen haben.	0	---	<i>outgroup</i>	29	59,2
Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst. ^c	0	---			
Menschen im Allgemeinen. ^d	29	59,2			
Fehlend	1	---	---	1	---

^a z.B. Familie, Freunde, Lebenspartner etc.

^b z.B. Nachbarn, Bekannte, Kollegen etc.

^c z.B. andere Religionen, andere Nationalitäten etc.

^d z.B. Fremde, die Ihnen zufällig auf der Straße begegnen

(Quelle: Eigene Berechnung)

Tab. 7.33: Vertrauensradius für Standard-Skala (PT₁) (N = 50)

Ausprägung	Häufigkeit	Gültige %	Ausprägung	Häufigkeit	Gültige %
Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben. ^a	4	8,2	<i>ingroup</i>	21	42,9
Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren. ^b	17	34,7			
Personen, die Sie erst ein einziges Mal getroffen haben.	0	---	<i>outgroup</i>	28	57,1
Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst. ^c	0	---			
Menschen im Allgemeinen. ^d	28	57,1			
Fehlend	1	---	---	1	---

^a z.B. Familie, Freunde, Lebenspartner etc.^b z.B. Nachbarn, Bekannte, Kollegen etc.^c z.B. andere Religionen, andere Nationalitäten etc.^d z.B. Fremde, die Ihnen zufällig auf der Straße begegnen

(Quelle: Eigene Berechnung)

Ausgehend von den absoluten und relativen Häufigkeiten ergibt sich zwischen Objektiv- und Standard-Skala kein Unterschied hinsichtlich der Anteile der verschiedenen Referenzgruppen. Innerhalb von *Versuchsgruppe 1* scheint die Verwendung der Objektiv-Skala somit keinen Beitrag zur Lösung des Radiusproblems zu leisten. Zwar gaben in beiden Fällen die Befragten mehrheitlich die *outgroup* als Vertrauensreferenz an, allerdings erscheint der Unterschied zum Anteil der *ingroup* verhältnismäßig gering.

Tab. 7.34: Konfidenzintervalle der Stichprobenkennwerte

<i>Objektiv-Skala (PR₁)</i> ($\sigma_{\bar{x}} = 0,07094$) ^a	
<i>ingroup</i>	Konf _{99%} = 0,2179 ≤ p ≤ 0,5984
<i>outgroup</i>	Konf _{99%} = 0,4016 ≤ p ≤ 0,7821
<i>Standard-Skala (PT₁)</i> ($\sigma_{\bar{x}} = 0,07143$) ^a	
<i>ingroup</i>	Konf _{99%} = 0,2370 ≤ p ≤ 0,6202
<i>outgroup</i>	Konf _{99%} = 0,3798 ≤ p ≤ 0,7630

^a Standardstichprobenfehler (Bortz (2005), S. 104)

(Quelle: Eigene Berechnung)

Dieser Eindruck wird auch durch die 99%-Konfidenzintervalle der Anteilswerte bestätigt. So überschneiden sich die Intervalle der wahren Anteile von *in-* und *outgroup* sowohl für die Objektiv- als auch für die Standard-Skala. Es ist daher anzunehmen, dass sich die wahren Anteilswerte der beiden Vertrauensreferenzen nicht unterscheiden bzw. sogar eine Umkehrung des Verhältnisses der Anteilswerte im Bereich des Möglichen liegt. Es muss somit davon ausgegangen werden, dass innerhalb von *Gruppe 1* weder die Standard- noch die Objektiv-Skala dazu in der Lage ist, eine exklusive Abbildung Generalisierten Vertrauens zu gewährleisten. Tatsächlich wird durch beide Skalen in fast der Hälfte der Fälle Partikulares Vertrauen abgebildet.

7.2.4.2 Gruppe 2: Standard/Objektiv

Der folgende Abschnitt behandelt den Vergleich des Vertrauensradius von Standard- und Objektiv-Skala in *Versuchsgruppe 2*

Tab. 7.35: Vertrauensradius für Standard-Skala (PR₂) (N = 50)

Ausprägung	Häufigkeit	Gültige %	Ausprägung	Häufigkeit	Gültige %
Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben. ^a	4	8,0	<i>ingroup</i>	16	32,0
Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren. ^b	12	24,0			
Personen, die Sie erst ein einziges Mal getroffen haben.	2	4,0	<i>outgroup</i>	34	68,0
Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst. ^c	0	---			
Menschen im Allgemeinen. ^d	32	64,0			
Fehlend	0	---	---	0	---

^a z.B. Familie, Freunde, Lebenspartner etc.^b z.B. Nachbarn, Bekannte, Kollegen etc.^c z.B. andere Religionen, andere Nationalitäten etc.^d z.B. Fremde, die Ihnen zufällig auf der Straße begegnen

(Quelle: Eigene Berechnung)

Tab. 7.36: Vertrauensradius für Objektiv-Skala (PT₂) (N = 50)

Ausprägung	Häufigkeit	Gültige %	Ausprägung	Häufigkeit	Gültige %
Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben. ^a	3	6,0	<i>ingroup</i>	14	28,0
Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren. ^b	11	22,0			
Personen, die Sie erst ein einziges Mal getroffen haben.	0	---	<i>outgroup</i>	36	72,0
Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst. ^c	1	2,0			
Menschen im Allgemeinen. ^d	35	70,0			
Fehlend	0	---	---	0	---

^a z.B. Familie, Freunde, Lebenspartner etc.^b z.B. Nachbarn, Bekannte, Kollegen etc.^c z.B. andere Religionen, andere Nationalitäten etc.^d z.B. Fremde, die Ihnen zufällig auf der Straße begegnen

(Quelle: Eigene Berechnung)

Auch innerhalb von *Versuchsgruppe 2* lässt sich, was die Relation des Anteils von *ingroup* zu *outgroup* als Vertrauensreferenz angeht, kein nennenswerter Unterschied zwischen Standard- und Objektiv-Skala feststellen. Somit scheint auch hier die Verwendung einer Objektiv-Skala keine Verbesserung hinsichtlich des Radiusproblems zu bewirken. Allerdings hat sich im Vergleich zu *Versuchsgruppe 1* der Anteil der Probanden, die die *outgroup* als Referenzgruppe des Vertrauens angaben, erhöht und die Differenz zwischen *in-* und *outgroup* ist größer geworden. Dies spricht dafür, dass in *Gruppe 2* eine größere Chance besteht, dass sowohl die Standard- als auch die Objektiv-Skala ausschließlich Generalisiertes Vertrauen abbildet.

Tab. 7.37: Konfidenzintervalle der Stichprobenkennwerte

Standard-Skala (PR₂) ($\sigma_x = 0,06664$)^a	
<i>ingroup</i>	Konf _{99%} = 0,1414 ≤ p ≤ 0,4519
<i>outgroup</i>	Konf _{99%} = 0,5014 ≤ p ≤ 0,8586
Objektiv-Skala (PT₂) ($\sigma_x = 0,06414$)^a	
<i>ingroup</i>	Konf _{99%} = 0,1081 ≤ p ≤ 0,4519
<i>outgroup</i>	Konf _{99%} = 0,5481 ≤ p ≤ 0,8919

^a Standardstichprobenfehler (Bortz (2005), S. 104)

(Quelle: Eigene Berechnung)

Betrachtet man die 99%-Konfidenzintervalle⁸⁰ der Anteilswerte wird sichtbar, dass sich die potentiellen Wertebereiche der wahren Anteile von *ingroup* und *outgroup* weder für die Standard-Skala noch für die Objektiv-Skala überschneiden. Es scheint somit ein bedeutsamer Unterschied zwischen den Anteilswerten der verschiedenen Referenzgruppen zu bestehen. Dabei ist der Anteil derjenigen, die die *outgroup* als Referenz angeben würden, jeweils deutlich größer als der Anteil derjenigen, die die *ingroup* als Referenz angeben würden. Von einer exklusiven Abbildung Generalisierten Vertrauens kann allerdings auch in *Versuchsgruppe 2* nicht gesprochen werden. Dazu fällt das Konfidenzintervall für den Anteil der *ingroup* zu breit aus.

7.2.4.3 Gruppe 3: Objektiv

Der nächste Absatz betrachtet das Verhältnis der Vertrauensreferenzen *ingroup* und *outgroup* für die Objektiv-Skala aus *Versuchsgruppe 3*.

Tab. 7.38: Vertrauensradius für Objektiv-Skala (PT₃) (N = 52)

Ausprägung	Häufigkeit	Gültige %	Ausprägung	Häufigkeit	Gültige %
Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben. ^a	2	3,8	<i>ingroup</i>	19	36,5
Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren. ^b	17	32,7			
Personen, die Sie erst ein einziges Mal getroffen haben.	1	1,9	<i>outgroup</i>	33	63,5
Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst. ^c	2	3,8			
Menschen im Allgemeinen. ^d	30	57,7			
Fehlend	0	---	---	0	---

^a z.B. Familie, Freunde, Lebenspartner etc.

^b z.B. Nachbarn, Bekannte, Kollegen etc.

^c z.B. andere Religionen, andere Nationalitäten etc.

^d z.B. Fremde, die Ihnen zufällig auf der Straße begegnen

(Quelle: Eigene Berechnung)

Die Verteilung der Probanden auf die Vertrauensreferenzen *ingroup* bzw. *outgroup* in *Versuchsgruppe 3* folgt in der Tendenz der Relation in *Versuchsgruppe 2*. Damit scheint die Objektiv-Skala eher Generalisiertes als Partikulares Vertrauen abzubilden.

Tab. 7.39: Konfidenzintervalle der Stichprobenkennwerte

Objektiv-Skala (PT ₃) ($\sigma_{\bar{x}} = 0,06743$) ^a	
<i>ingroup</i>	Konf _{99%} = 0,1850 ≤ p ≤ 0,5458
<i>outgroup</i>	Konf _{99%} = 0,4542 ≤ p ≤ 0,8150

^a Standardstichprobenfehler (Bortz (2005), S. 104)

(Quelle: Eigene Berechnung)

Betrachtet man allerdings die 99%-Konfidenzintervalle für die Anteilswerte der beiden Vertrauensreferenzen wird deutlich, dass sich die wahren Anteilswerte von *ingroup* und

⁸⁰ Die Berechnung des 99%-Konfidenzintervalls mit Hilfe einer Normalverteilungsapproximation der Binomialverteilung ist opportun, da gilt $0,205 * 0,795 * 190 = 30,97 > 9$ ($p * q * n > 9$). Die Stichprobe ist damit ausreichend groß für die Schätzung des Standardfehlers und der Intervallbreite. (Bortz (2005), S. 78, S. 104)

7. Hauptstudie

outgroup kaum unterscheiden. Aus der Überlappung der Wertebereiche könnte tatsächlich sogar ein umgekehrtes Anteilsverhältnis resultieren. Auch durch eine singuläre Präsentation der Objektiv-Skala ohne potentiell Priming durch eine parallele Vertrauensmessung per Standard-Skala ergibt sich somit keine Lösung für das Radiusproblem. Vielmehr bestätigt sich das bereits im Rahmen der Pilotstudie gefundene Ergebnis.

7.2.4.4 Gruppe 4: Standard

Der nächste Absatz betrachtet das Verhältnis der Vertrauensreferenzen *ingroup* und *outgroup* für die Standard-Skala aus *Versuchsgruppe 4*.

Tab. 7.40: Vertrauensradius für Standard-Skala (PT₄) (N = 52)

Ausprägung	Häufigkeit	Gültige %	Ausprägung	Häufigkeit	Gültige %
Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben. ^a	7	13,5	<i>ingroup</i>	20	38,5
Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren. ^b	13	25,0			
Personen, die Sie erst ein einziges Mal getroffen haben.	0	---	<i>outgroup</i>	32	61,5
Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst. ^c	0	---			
Menschen im Allgemeinen. ^d	32	61,5			
Fehlend	0	---	---	---	---

^a z.B. Familie, Freunde, Lebenspartner etc.

^b z.B. Nachbarn, Bekannte, Kollegen etc.

^c z.B. andere Religionen, andere Nationalitäten etc.

^d z.B. Fremde, die Ihnen zufällig auf der Straße begegnen

(Quelle: Eigene Berechnung)

Das Resultat der Radiusanalyse in *Versuchsgruppe 4* bestätigt die fortdauernde Existenz des Radiusproblems innerhalb der Vertrauensmessung. Zwar nennt die Mehrzahl der Probanden die *outgroup* als Referenz für das angegebene Vertrauen, gleichzeitig ist aber auch die Anzahl der Probanden, die bekannte Personen als Referenzgruppe angeben verhältnismäßig hoch.

Tab. 7.41: Konfidenzintervalle der Stichprobenkennwerte

Standard-Skala (PT ₄) ($\sigma_{\bar{x}} = 0,06812$) ^a	
<i>ingroup</i>	Konf _{99%} = 0,2023 ≤ p ≤ 0,5669
<i>outgroup</i>	Konf _{99%} = 0,4331 ≤ p ≤ 0,7977

^a Standardstichprobenfehler (Bortz (2005), S. 104)

(Quelle: Eigene Berechnung)

Tatsächlich zeigt sich, dass sich die wahren Anteile der beiden Referenzgruppen kaum voneinander unterscheiden. Es kann somit nach wie vor bezweifelt werden, dass die Standard-Skala ausschließlich Generalisiertes Vertrauen misst. Vielmehr ist zu erwarten, dass in bis zu der Hälfte der Fälle Partikulares Vertrauen abgebildet wird.

7.2.4.5 Zusammenfassung

Vergleicht man die Objektiv- und Standard-Skalen über alle Versuchsgruppen hinweg, fällt auf, dass der Anteil der Probanden, die die *outgroup* als Vertrauensreferenz angeben, für die Objektiv-Skala grundsätzlich höher ist als für die Standard-Skala. Dies stellt einen Hinweis darauf

dar, dass die Objektiv-Skala in der Tendenz besser zwischen Partikularem und Generalisiertem Vertrauen trennen kann bzw. besser geeignet ist für eine exklusive Abbildung Generalisierten Vertrauens als die Standard-Skala. Einschränkend ist allerdings zu bemerken, dass der Unterschied jeweils vergleichsweise gering ausfällt.

Mit Ausnahme der Skalen in *Versuchsgruppe 2* überschneiden sich für alle Skalen die wahren Anteile von *ingroup* und *outgroup*. Es scheint somit in Wahrheit kein großer Unterschied zwischen der Anzahl der Personen, die ihnen bekannte Personen als Vertrauensreferenz angeben, und der Anzahl von Personen, die generalisierte Andere als Vertrauensreferenz angeben, zu bestehen. Zwar lässt sich die oben angesprochene Tendenz zur Lösung des Radiusproblems durch die Objektiv-Skala entdecken, insgesamt scheinen sich die Objektiv- und Standard-Skala aber nicht nachhaltig in ihrer Differenzierungsfähigkeit hinsichtlich Partikularem und Generalisiertem Vertrauen zu unterscheiden.

7.2.5 Prüfung auf Normalverteilung⁸¹

Um ein geeignetes Testverfahren für die Differenz zwischen den verschiedenen Messmethoden zur Erfassung Generalisierten Vertrauens zu ermitteln bzw. um zu bestimmen, ob für die Prüfung des Unterschieds zwischen Objektiv- und Standard-Messung auf Signifikanz parametrische oder nonparametrische Verfahren angemessen sind, sollen zunächst die Summenscores der verschiedenen Vertrauensskalen in jeder der Experimentalgruppen auf Normalverteilung hin untersucht werden.

Es ergeben sich dabei die folgenden Hypothesen:

H₀: Die Verteilungsfunktion des Summenscores über die Standard-Skala bzw. die Objektiv-Skala bzw. die Objektiv-Kurzskala weicht in der Population nicht von der theoretisch erwarteten Verteilungsform (Normalverteilung) ab. Keine Ausprägung der genannten Summenscores weicht von der erwarteten Verteilungsform (Normalverteilung) ab.

$$\Phi(\text{Standard}) = \Phi_0(\text{Standard}) \text{ für alle } x$$

$$\Phi(\text{Objektiv}) = \Phi_0(\text{Objektiv}) \text{ für alle } x$$

$$\Phi(\text{Objektiv [Kurz]}) = \Phi_0(\text{Objektiv [Kurz]}) \text{ für alle } x$$

H₁: Die Verteilungsfunktion des Summenscores über die Standard-Skala bzw. die Objektiv-Skala bzw. die Objektiv-Kurzskala entspricht innerhalb der Population nicht der theoretisch erwarteten Verteilungsform (Normalverteilung). Mindestens eine Ausprägung der genannten Summenscores weicht von der erwarteten Verteilungsform (Normalverteilung) ab.

$$\Phi(\text{Standard}) \neq \Phi_0(\text{Standard}) \text{ für alle } x$$

$$\Phi(\text{Objektiv}) \neq \Phi_0(\text{Objektiv}) \text{ für alle } x$$

$$\Phi(\text{Objektiv [Kurz]}) \neq \Phi_0(\text{Objektiv [Kurz]}) \text{ für alle } x$$

⁸¹ Vgl. Anhang II: Exkurse Signifikanzprüfungen, II.a Exkurs: Prüfung auf Normalverteilung

Tab. 7.42: Kolmogorov-Smirnov-Anpassungstest mit Lilliefors-Korrektur

Gruppe 1: Objektiv/Standard	
Standard-Skala	$D_{\max} = ,249$; $Z = 1,762$; (zweiseitig); $p < ,01$
Objektiv-Skala	$D_{\max} = ,391$; $Z = 2,766$; (zweiseitig); $p < ,01$
Objektiv-Kurzskala	$D_{\max} = ,300$; $Z = 2,121$; (zweiseitig); $p < ,01$
Gruppe 2: Standard/Objektiv	
Standard-Skala	$D_{\max} = ,245$; $Z = 1,732$; (zweiseitig); $p < ,01$
Objektiv-Skala	$D_{\max} = ,392$; $Z = 2,774$; (zweiseitig); $p < ,01$
Objektiv-Kurzskala	$D_{\max} = ,383$; $Z = 2,710$; (zweiseitig); $p < ,01$
Gruppe 3: Objektiv	
Objektiv-Skala	$D_{\max} = ,320$; $Z = 2,309$; (zweiseitig); $p < ,01$
Objektiv-Kurzskala	$D_{\max} = ,340$; $Z = 2,455$; (zweiseitig); $p < ,01$
Gruppe 4: Standard	
Standard-Skala	$D_{\max} = ,242$; $Z = 1,747$; (zweiseitig); $p < ,01$

D_{\max} : Prüfgröße bzw. Maximum aller Abweichungsbeträge zwischen empirischer und theoretischer Verteilungsfunktion
 Z : Kolmogorov-Smirnov-Z

Betrachtet man das Ergebnis des Kolmogorov-Smirnov-Anpassungstests für *Gruppe 1*, so zeigt sich, dass die Verteilungsfunktionen sämtlicher betrachteten Skalen-Summenscores hochsignifikant von einer Normalverteilung abweichen. Mit anderen Worten: Die Summe der Abweichungsbeträge zwischen empirischer und theoretischer Verteilungsfunktion liegen oberhalb des kritischen Wertes für sowohl das 5%-Signifikanzniveau als auch für das 1%-Signifikanzniveau. Es wird daher die H_1 beibehalten und die H_0 abgelehnt. Dieses Ergebnis zeigt außerdem an, dass zur Prüfung von Unterschiedshypothesen zwischen den verschiedenen Skalen innerhalb von *Gruppe 1* nonparametrische Verfahren verwendet werden sollten.

Auch innerhalb von *Gruppe 2* ergeben sich für jede der betrachteten Skalen hochsignifikante Abweichungen zwischen empirischer und theoretischer Verteilungsfunktion, sodass jeweils die H_0 abgelehnt und die H_1 beibehalten werden muss. Entsprechend dem Kolmogorov-Smirnov-Test ist demnach weder für die Standard-Skala noch für die Objektiv-Skala bzw. -Kurzskala Normalverteilung gegeben und es sollten innerhalb von *Gruppe 2* nonparametrische Verfahren zur Prüfung von Unterschiedshypothesen zwischen den einzelnen Vertrauensskalen verwendet werden.

Der Kolmogorov-Smirnov-Test ergibt für *Gruppe 3*, dass sich die sowohl die Verteilung des Summenscores der Objektiv-Skala als auch die Verteilung des Summenscores der Objektiv-Kurzskala hochsignifikant von einer Normalverteilung unterscheiden. Die Prüfgröße überschreitet für beide Skalen den kritischen Wert für $\alpha=1\%$. Es wird demnach die H_0 abgelehnt und die H_1 beibehalten. Zur Prüfung von Unterschiedshypothesen, deren Gegenstand die Summenscores der beiden überprüften Skalen sind, sollten nonparametrische Verfahren benutzt werden.

Die Auswertung des Kolmogorov-Smirnov-Anpassungstest ergibt für *Gruppe 4* einen hochsignifikanten Unterschied zwischen empirischer und erwarteter Verteilungsform. Entsprechend wird die H_0 abgelehnt und die H_1 beibehalten. Da für die Standard-Skala in *Gruppe 4* keine Normalverteilung angenommen werden kann, sollten zur Prüfung von Unterschiedshypothesen nonparametrische Verfahren verwendet werden.

In keiner der Versuchsgruppen und für keine der betrachteten Skalen konnte bei der Prüfung auf Normalverteilung die Nullhypothese beibehalten werden. Mit anderen Worten: Die Verteilungen der Summenscores in den voneinander unabhängigen Vertrauensmessungen weichen jeweils hochsignifikant von der erwarteten Verteilungsfunktion (hier: Normalverteilung) ab. Es sollten daher sowohl für die Prüfung von Unterschiedshypothesen

innerhalb der Gruppen als auch für deren Prüfung über die Gruppengrenzen hinweg nonparametrische Verfahren angewendet werden.

7.2.6 Unterschiedshypothesen und Mittelwertvergleiche

Handlungsleitend für die nachfolgenden Analysen ist die Annahme, dass die Erfassung von Generalisiertem Vertrauen mit Hilfe der Standard-Variante eine Überschätzung des wahren Wertes Generalisierten Vertrauens bewirkt. Die Erfassung Generalisierten Vertrauens auf Basis der Objektiv-Variante stellt dagegen eine validere Abbildung des wahren Wertes Generalisierten Vertrauens dar. Entsprechend sollten die Probanden auf der Objektiv-Skala bzw. -Kurzskala ein signifikant niedrigeres Generalisiertes Vertrauen angeben als auf der Standard-Skala.

7.2.6.1 Exkurs: Parametrische Verfahren⁸²

Als Grundvoraussetzung für die Verwendung von parametrischen Verfahren gilt im Allgemeinen, dass das zu untersuchende Merkmal in den Populationen, aus denen die in die Signifikanzprüfung eingehenden Stichproben stammen, normalverteilt ist. Wie im vorangegangenen Abschnitt gezeigt werden konnte, kann diese Vorgabe auf Basis der erhobenen Daten unabhängig von der Versuchsgruppe für keine der Vertrauensmessungen als erfüllt angesehen werden. Es kann allerdings angenommen werden, dass parametrische Verfahren für „große“ Stichproben ($N > 30$) auch dann valide Ergebnisse liefern, wenn das zu untersuchende Merkmal nicht normalverteilt ist. (Bortz (2005), S. 141; Bortz/Lienert/Boehnke (2008), S. 81; Bühner/Ziegler (2009), S. 270; Kubinger/Rasch/Moder (2009), S. 26; Rasch/Guiard (2004), S. 175)

Eine weitere Voraussetzung parametrischer Verfahren, die insbesondere den Vergleich unabhängiger Stichproben betrifft, stellt die Notwendigkeit homogener Varianzen in den zugrundeliegenden Populationen dar. Daneben können aber auch Varianzunterschiede in den zu vergleichenden Stichproben das Ergebnis der Signifikanzprüfung sowohl progressiv als auch konservativ beeinflussen. Im Falle von kleinen Stichproben ($N \leq 30$) wird daher empfohlen, grundsätzlich auf Varianten parametrischer Verfahren zurückzugreifen, die eine Heterogenität der Varianzen annehmen. (Bortz (2005), S. 141; Brosius (2011), S. 488; Eid/Gollwitzer/Schmitt (2015), S. 336; Kubinger/Rasch/Moder (2009), S. 26)

Generell gilt, dass parametrische Verfahren relativ robust gegenüber Voraussetzungsverletzungen reagieren bzw. erst drastische Abweichungen von obengenannten Forderungen eine Ergebnisverzerrung bewirken. Zudem weisen sie in der Regel eine höhere Teststärke auf als nonparametrische Verfahren. Sind die Voraussetzungen für parametrische Verfahren verletzt, so schlagen BORTZ/LIENERT/BOEHNKE eine Parallelprüfung mit Hilfe von verteilungsfreien Verfahren vor. Bei einem gleichlautenden Ergebnis kann das Resultat des parametrischen Tests als gültig angenommen werden. Bei einem entgegengesetzten Ergebnis bekommt das Resultat des nonparametrischen Tests den Vorzug. (Bortz (2005), 141ff; Bortz/Lienert/Boehnke (2008), S. 81, S. 85f; Bühner/Ziegler (2009), S. 270; Pospeschill (2006), S. 386; Sedlmeier/Renkewitz (2008), S. 581; Siegel/Castellan (1988), S. 36)

⁸² Vgl. Anhang II: Exkurse Signifikanzprüfungen, II.b Exkurs: Parametrische Verfahren

Im Angesicht der augenscheinlichen Verletzungen der Voraussetzungen parametrischer Verfahren bei gleichzeitig zu erwartender höherer Teststärke im Vergleich zu verteilungsfreien Verfahren, soll in den nachfolgenden Analyseschritten der Empfehlung von BORTZ/LIENERT/BOEHNKE gefolgt werden. So werden zunächst Mittelwertvergleiche anhand parametrischer Verfahren durchgeführt, die im Anschluss mit Hilfe von nonparametrischen Verfahren validiert werden.

7.2.6.2 Exkurs: Nonparametrische Verfahren⁸³

Ausgehend von den Ergebnissen des Kolmogorov-Smirnov-Anpassungstests darf für die Verteilungen der Summenscores der Standard-Skala und der Objektiv-Skala bzw. -Kurzskala keine Normalverteilung angenommen werden. Um zu prüfen, ob sich das Antwortverhalten der Probanden zwischen der Standard-Variante der Vertrauensmessung und der Objektiv-Variante der Vertrauensmessung unterscheidet, kommen daher sogenannte nonparametrische Verfahren zum Einsatz. Solche Prüfverfahren werden auch als verteilungsfreie Verfahren bezeichnet und setzen keine spezielle Verteilungsannahme voraus. (Eid/Gollwitzer/Schmitt (2015), S. 308; Bühner/Ziegler (2009), S. 268f; Siegel/Castellan (1988), S. 34)

Die Verletzung der Normalverteilungsvoraussetzung kann auch durch eine vergleichsweise geringe Stichprobengröße bedingt sein. Entsprechend stellt die geringe Fallzahl innerhalb der Versuchsgruppen einen weiteren Grund für die Verwendung nonparametrischer Verfahren dar. (Sedlmeier/Renkewitz (2013), S. 18; Siegel/Castellan (1988), S. 35)

7.2.6.3 Kontrolle von Effekten der experimentellen Situation⁸⁴

Die Gültigkeit etwaiger Unterschiede zwischen den Objektiv-Skalen und der Standard-Skala kann durch sogenannte *Pretest*-, *Positions*- oder *Carry-over*-Effekte bedroht sein. Um sicher zu stellen, dass allein das Treatment, d.h. die Variation des Messverfahrens, für die Unterschiede im Antwortverhalten zwischen Objektiv- und Standard-Skala verantwortlich ist, werden nachfolgend verschiedene Kontrollprozeduren angewendet. Hierbei handelt es sich im Wesentlichen um Prüfungen auf signifikante Unterschiede zwischen Bestandteilen des Surveyexperimentes, die aufgrund ihres Expositionszeitpunktes keinesfalls durch eine der obengenannten Verzerrungen beeinflusst sein können (PR_1 , PR_2 , PT_3 und PT_4), und solchen Bestandteilen, die aufgrund ihrer relativen Stellung zu den anderen Bestandteilen des Surveyexperimentes durch die Auswirkungen von Vorher-Messungen in ihrer Validität beeinträchtigt sein können (PT_1 und PT_2). Es ergeben sich aus dieser Überlegung dann die folgenden Paarungen PR_1 vs. PT_2 , PR_2 vs. PT_1 , PT_1 vs. PT_4 und PT_2 vs. PT_3 , wobei Verzerrungen durch die genannten Störeffekte dann als nicht existent angenommen werden können, wenn sich für keine der Signifikanzprüfungen eine signifikante Differenz im Antwortverhalten ergibt. Da es sich bei den Testpaarungen um eine Anordnung als *between-subject*-Design handelt, d.h. voneinander unabhängige Stichproben gegeneinander getestet werden, kommt für sämtliche nachfolgende Signifikanzprüfungen als parametrisches Verfahren die einfaktorielle Varianzanalyse und als nonparametrisches Verfahren der Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben zum Einsatz.

⁸³ Vgl. Anhang II: Exkurse Signifikanzprüfungen, II.c Exkurs: Nonparametrische Verfahren

⁸⁴ Vgl. Abschnitt „7.1.1.5.1 Kontrolltechniken“

Eine weitere rudimentäre Prüfung auf Biasfreiheit erfolgt im Rahmen der Gegenüberstellung von PR_1 vs. PT_1 und PR_2 vs. PT_2 im Kapitel „Treatment-Effekt“. Zur Kontrolle von Positionseffekten wurden in den beiden *Versuchsgruppen 1* und *2* jeweils invers zueinander verlaufende Abfolgen von Objektiv-Skala und Standard-Skala realisiert. Ergeben sich nun beim Vergleich von Pretest und Posttest innerhalb der einzelnen Gruppen für beide Gruppen das gleiche Ergebnis, so kann angenommen werden, dass die Reihenfolge der Bearbeitung von Objektiv-Skala und Standard-Skala durch die Probanden keinen Einfluss auf deren Antwortverhalten hat. Vielmehr können Unterschiede im Antwortverhalten auf das Treatment zurückgeführt werden.

Vorgreifend auf das Kapitel „Treatment-Effekt“ sei bereits an dieser Stelle berichtet, dass sich im Zuge der Signifikanzprüfungen von PR_1 vs. PT_1 und PR_2 vs. PT_2 jeweils ein signifikanter Unterschied im Antwortverhalten zwischen Objektiv- und Standard-Skala ergeben hat. Gleiches gilt für den Vergleich von Objektiv-Kurzskala und Standard-Skala. Zudem zeigte der Paarvergleich für alle Unterschiedsprüfungen, dass die Mehrzahl der Fälle auf der Objektiv-Variante der Vertrauensmessung niedrigere Werte erreichten als auf der Standard-Variante. Somit stimmt auch die Richtung der festgestellten signifikanten Unterschiede in den beiden Gruppen und über alle Testpaarungen hinweg überein. Hieraus kann geschlossen werden, dass die Probanden in keiner der beiden Versuchsgruppen durch den jeweiligen Pretest ein Priming erfahren haben, welches zu einem wechselseitig konsistenten Antwortverhalten auf der Objektiv- bzw. Standard-Skala geführt hat. Dieses Ergebnis stellt zudem einen Hinweis darauf dar, dass die Verschleierung der Messintention auf der Objektiv-Skala mittels Items, die eine geringe Augenscheinvalidität aufweisen, erfolgreich ist.

7.2.6.4 Parametrische Signifikanzprüfung

Zunächst soll die Prüfung auf Effekte der experimentellen Situation anhand von parametrischen Verfahren vollzogen werden. Aus Gründen der Testökonomie wurde hierzu die einfaktorielle Varianzanalyse als Prüfverfahren eingesetzt. So können, im Gegensatz zum T-Test für unabhängige Stichproben, mit Hilfe der einfaktoriellen Varianzanalyse die notwendigen Signifikanzprüfungen zwischen den Objektiv-Skalen, den Objektiv-Kurzskalen und den Standard-Skalen in jeweils einer Analyse zusammengefasst werden, anstatt insgesamt sechs unterschiedliche Tests durchführen zu müssen.

7.2.6.4.1 Objektiv-Skalen: PR_1 vs. PT_2 vs. PT_3

Mit PR_1 , PT_2 und PT_3 werden jeweils per Objektiv-Skala durchgeführte Messungen Generalisierten Vertrauens gegenübergestellt. Dabei sind sowohl PR_1 als auch PT_3 unbeeinflusst durch ein Pretest-Priming, da in beiden Fällen keine Vertrauensmessung vorangestellt war. PT_2 war dagegen eine Messung Generalisierten Vertrauens per Standard-Skala vorangegangen. Es besteht somit die Möglichkeit, dass das Ergebnis der Objektiv-Skala aufgrund der Vorher-Messung verfälscht sein könnte. Ergibt sich kein signifikanter Unterschied im Antwortverhalten zwischen den genannten Gruppen, so kann davon ausgegangen werden, dass das Antwortverhalten in PT_2 dem Antwortverhalten in PR_1 und PT_3 entspricht und somit keine Beeinflussung von PT_2 durch PR_2 stattgefunden hat. Aus dieser Überlegung ergeben sich die folgenden Hypothesen:

7. Hauptstudie

H_0 : Die Gruppen PR_1 , PT_2 und PT_3 unterscheiden sich nicht hinsichtlich des Mittelwertes der Objektiv-Skala zur Erfassung Generalisierten Vertrauens und entstammen daher derselben Population.

$$\mu_{PR1} = \mu_{PT2} = \mu_{PT3}$$

H_1 : Mindestens zwei der Gruppen PR_1 , PT_2 und PT_3 unterscheiden sich hinsichtlich des Mittelwertes der Objektiv-Skala zur Erfassung Generalisierten Vertrauens und entstammen daher verschiedenen Populationen.

$$\mu_{PR1} \neq \mu_{PT2} \vee \mu_{PR1} \neq \mu_{PT3} \vee \mu_{PT2} \neq \mu_{PT3}$$

Tab. 7.43: Deskriptive Statistik und Prüfung auf Varianzhomogenität

Gruppe	Deskriptive Statistik			Levene-Test auf Varianzhomogenität	
	\bar{x}	s	N	F	Signifikanz
PR_1	2,8800	,52060	50	,900	,409 ^{n.s.}
PT_2	2,9600	,49322	50		
PT_3	3,0192	,61006	52		
Gesamt	2,9539	,54395	152		

^{n.s.} nicht signifikant ($p > .05$ (zweiseitig))

(Quelle: eigene Berechnungen)

Zunächst lässt sich feststellen, dass die Fehlervarianzen innerhalb der Gruppen nicht signifikant voneinander abweichen. Es kann somit Varianzhomogenität angenommen werden und die gemittelte Streuung als guter Schätzwert der Populationsvarianz angesehen werden. Ausgehend hiervon scheinen die Gruppenmittelwerte gleich gute Schätzungen des Populationsmittelwertes darzustellen. (Bühner/Ziegler (2009), S. 362f, S. 372ff, S. 378f)

Tab. 7.44: ANOVA (PR_1 vs. PT_2 vs. PT_3)

	QS ^a	df ^b	σ^2 ^c
Gruppe^d	,497	2	,248
Fehler	44,181	149	,297
Total	1371,000	152	---
F^e	,838		
Signifikanz	,435 ^{n.s.}		
η^2 partial^f	,011		
f^g	,105		
1-β	,192 ^h		
β	,808 ⁱ		

^a Quadratsumme

^b Freiheitsgrade („degrees of freedom“): $df_{treat} = k-1$; $df_{fehler} = k*(n-1)$; $df_{tot} = n*k-1$ mit k =Anzahl der Faktorstufen; n =Fallzahl (Bortz (2005), S. 251ff)

^c Mittel der Quadrate

^d Treatment

^e F-verteilte Prüfgröße

^f Effektstärke bzw. Anteil aufgeklärter Varianz an der Gesamtvarianz: $\eta^2_{partial} = \frac{QS_{treat}}{QS_{treat} + QS_{fehler}}$; $\eta^2 \approx .01 \rightarrow$ „kleiner“ Effekt, $\eta^2 \approx .06 \rightarrow$ „mittlerer“ Effekt, $\eta^2 \approx .14 \rightarrow$ „großer“ Effekt (Eid/Gollwitzer/Schmitt (2015), S. 409f; Bortz (2005), S. 255, S. 259, S. 280; Bühner/Ziegler (2009), S. 366ff)

^g Effektstärke: $f = \sqrt{\frac{\eta^2}{1-\eta^2}}$; $f = .10 \rightarrow$ kleiner Effekt; $f = .25 \rightarrow$ moderater Effekt; $f = .40 \rightarrow$ großer Effekt (Bühner/Ziegler (2009), S. 368; Cohen (1988), S. 284ff; Eid/Gollwitzer/Schmitt (2015), S. 411)

^h Teststärke bzw. beobachtete Schärfe für $\alpha=5\%$, $df_1=2$ u. $df_2=149$ (Bühner/Ziegler (2009), S. 370, S. 381)

ⁱ β -Fehler-Wahrscheinlichkeit für $\alpha=5\%$, $df_1=2$; $df_2=149$: „Die Irrtumswahrscheinlichkeit β ist die Wahrscheinlichkeit, mit der ein statistischer Test ein nicht-signifikantes Ergebnis ergibt, obwohl in Wirklichkeit die H_1 gilt.“ (Eid/Gollwitzer/Schmitt (2015), S. 224)

^{n.s.} nicht signifikant ($p > .05$ (zweiseitig))

(Quelle: eigene Berechnungen)

Ein Vergleich von QS_{treat} und QS_{fehler} legt nahe, dass Abweichungen vom Populationsmittelwert vornehmlich auf individuelle Variationen innerhalb der Gruppen zurückzuführen sind, aber kaum Unterschiede zwischen den Gruppen vorliegen. Die Abweichungen innerhalb der Gruppen leisten einen deutlich höheren Beitrag zur Gesamtvariation als die Abweichungen, die auf

7. Hauptstudie

Unterschiede zwischen den Gruppen zurückzuführen sind. (Bühner/Ziegler (2009), S. 342f; Eid/Gollwitzer/Schmitt (2015), S. 396f; Sedlmeier/Renkewitz (2008), S. 443ff; Sedlmeier/Renkewitz (2013), S. 437f)

Zusätzlich ist festzustellen, dass die Varianz zwischen den Gruppen leicht niedriger ausfällt als die Varianz innerhalb der Gruppen bzw. beide nahezu identisch sind. Es kann daher angenommen werden, dass durch die Gruppenzugehörigkeit der Probanden kein Effekt auf das Antwortverhalten ausgelöst wurde. (Bühner/Ziegler (2009), S. 335; Eid/Gollwitzer/Schmitt (2015), S. 404)

Wie aufgrund der bisherigen Interpretation zu erwarten, zeigt die F-verteilte Prüfgröße ein nicht signifikantes Ergebnis an. Etwaige Unterschiede zwischen den Gruppen folgen demnach keiner Systematik und es sollte die H_0 beibehalten sowie die H_1 abgelehnt werden.

Die Gruppenzugehörigkeit erklärt mit 1,1% ($\eta^2_{\text{partial}} = .011$) nur einen geringen Anteil der Variation in der abhängigen Variable „Generalisiertes Vertrauen“. Auch die Effektstärke zeigt mit $f = .105$ nur einen geringen Kontrast zwischen den Versuchsgruppen an. (Bühner/Ziegler (2009), S. 366ff; Cohen (1988), S. 284ff; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 194f; Sedlmeier/Renkewitz (2008), S. 452ff)

Die post-hoc-Teststärke erreicht mit $1 - \beta = .192$ nicht das erforderliche Niveau von 80% für eine statistische Absicherung des Ergebnisses. Die Wahrscheinlichkeit zu Gunsten einer korrekten H_1 zu entscheiden ist damit erheblich reduziert. Gleichzeitig ist mit $\beta = .808$ die Gefahr eines β -Fehlers stark erhöht ($\beta > .20$). Möglicherweise ist die Stichprobengröße zu gering, als dass die H_1 angenommen werden könnte. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff, S. 369ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2008), S. 455ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 446ff)

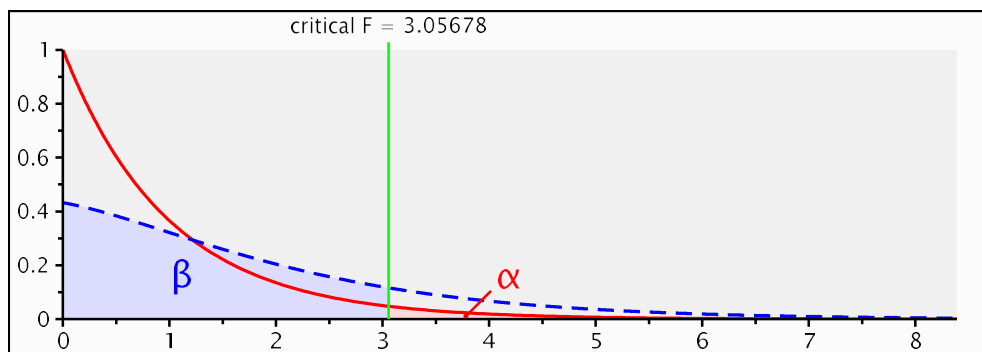


Abb. 7.10: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PR_1 vs. PT_2 vs. PT_3 bei einseitiger Testung (ANOVA) und $\alpha = 5\%$, $df_1 = 2$, $df_2 = 149$

7.2.6.4.2 Objektiv-Kurzskalen: $PR_{1(KS)}$ vs. $PT_{2(KS)}$ vs. $PT_{3(KS)}$

Im folgenden Abschnitt werden die aus PR_1 , PT_2 und PT_3 hervorgehenden Objektiv-Kurzskalen gegeneinander getestet. Hierbei gelten die gleichen Prämissen wie schon bei der Kontrastierung der Objektiv-Skalen. So ist für PR_1 und PT_3 anzunehmen, dass kein Pretest-, Positions- oder Carry-over-Effekt die Vertrauensmessung beeinflusst hat. Um auszuschließen, dass PT_2 durch ein Pretest-Priming verzerrt worden sein könnte, sollten sich die Mittelwerte der Gruppen nicht voneinander unterscheiden. Es werden deshalb die nachfolgenden Hypothesen überprüft:

7. Hauptstudie

H_0 : Die Gruppen PR_1 , PT_2 und PT_3 unterscheiden sich nicht hinsichtlich des Mittelwertes der Objektiv-Kurzskala zur Erfassung Generalisierten Vertrauens und entstammen daher derselben Population.

$$\mu_{PR1(KS)} = \mu_{PT2(KS)} = \mu_{PT3(KS)}$$

H_1 : Mindestens zwei der Gruppen PR_1 , PT_2 und PT_3 unterscheiden sich hinsichtlich des Mittelwertes der Objektiv-Kurzskala zur Erfassung Generalisierten Vertrauens und entstammen daher verschiedenen Populationen.

$$\mu_{PR1(KS)} \neq \mu_{PT2(KS)} \vee \mu_{PR1(KS)} \neq \mu_{PT3(KS)} \vee \mu_{PT2(KS)} \neq \mu_{PT3(KS)}$$

Tab. 7.45: Deskriptive Statistik und Prüfung auf Varianzhomogenität

Gruppe	Deskriptive Statistik			Levene-Test auf Varianzhomogenität	
	\bar{x}	s	N	F	Signifikanz
$PR_{1(KS)}$	3,0000	,63888	50	,031	,970 ^{n.s.}
$PT_{2(KS)}$	3,1400	,53490	50		
$PT_{3(KS)}$	3,0769	,58899	52		
Gesamt	3,0724	,58798	152		

^{n.s.} nicht signifikant ($p > .05$ (zweiseitig))

(Quelle: eigene Berechnungen)

Die Fehlervarianzen der einzelnen Gruppen weichen nicht signifikant voneinander ab, sodass Varianzhomogenität angenommen werden kann. Die über alle Gruppen hinweg gemittelte Streuung kann somit als erwartungstreuer Schätzwert für die Populationsvarianz angenommen werden. Entsprechend gelten auch die Gruppenmittelwerte als gute Schätzung des Populationsmittelwertes. (Bühner/Ziegler (2009), S. 362f, S. 372ff, S. 378f)

Tab. 7.46: ANOVA ($PR_{1(KS)}$ vs. $PT_{2(KS)}$ vs. $PT_{3(KS)}$)

	QS ^a	df ^b	σ^2 ^c
Gruppe^d	,492	2	,246
Fehler	51,712	149	,347
Total	1487,000	152	---
F^e		,708	
Signifikanz		,494 ^{n.s.}	
η^2 partial^f		,009	
f^g		,095	
1-β		,168 ^h	
β		,832 ⁱ	

^a Quadratsumme

^b Freiheitsgrade („degrees of freedom“): $df_{treat} = k - 1$; $df_{fehler} = k * (n - 1)$; $df_{tot} = n * k - 1$ mit k =Anzahl der Faktorstufen; n =Fallzahl (Bortz (2005), S. 251ff)

^c Mittel der Quadrate

^d Treatment

^e F-verteilte Prüfgröße

^f Effektstärke bzw. Anteil aufgeklärter Varianz an der Gesamtvarianz: $\eta^2_{partial} = \frac{QS_{treat}}{QS_{treat} + QS_{fehler}}$; $\eta^2 \approx .01 \rightarrow$ „kleiner“ Effekt, $\eta^2 \approx .06 \rightarrow$ „mittlerer“

Effekt, $\eta^2 \approx .14 \rightarrow$ „großer“ Effekt (Eid/Gollwitzer/Schmitt (2015), S. 409f; Bortz (2005), S. 255, S. 259, S. 280; Bühner/Ziegler (2009), S. 366ff)

^g Effektstärke: $f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$; $f = .10 \rightarrow$ kleiner Effekt; $f = .25 \rightarrow$ moderater Effekt; $f = .40 \rightarrow$ großer Effekt (Bühner/Ziegler (2009),

S. 368; Cohen (1988), S. 284ff; Eid/Gollwitzer/Schmitt (2015), S. 411)

^h Teststärke bzw. beobachtete Schärfe für $\alpha = 5\%$, $df_1 = 2$ u. $df_2 = 149$ (Bühner/Ziegler (2009), S. 370, S. 381)

ⁱ $1 - \beta$ -Fehler-Wahrscheinlichkeit für $\alpha = 5\%$, $df_1 = 5$; $df_2 = 149$: „Die Irrtumswahrscheinlichkeit β ist die Wahrscheinlichkeit, mit der ein statistischer Test ein nicht-signifikantes Ergebnis ergibt, obwohl in Wirklichkeit die H_1 gilt.“ (Eid/Gollwitzer/Schmitt (2015), S. 224)

^{n.s.} nicht signifikant ($p > .05$ (zweiseitig))

(Quelle: eigene Berechnungen)

Aus einem Vergleich der Quadratsummen geht hervor, dass Abweichungen von der zentralen Tendenz der Grundgesamtheit größtenteils auf die Varianz innerhalb der Gruppen attribuiert werden kann. Es existieren kaum Abweichungen zwischen den Gruppen. Mit anderen Worten: Die Variation innerhalb der Gruppen hat einen deutlich höheren Anteil an der Gesamtvarianz als

die Fluktuationen, die auf Unterschiede zwischen den Gruppen zurückgehen. (Bühner/Ziegler (2009), S. 342f; Eid/Gollwitzer/Schmitt (2015), S. 396f; Sedlmeier/Renkewitz (2008), S. 443ff; Sedlmeier/Renkewitz (2013), S. 437f)

Die Varianz zwischen den Gruppen fällt geringer aus als die Varianz innerhalb der Gruppen. Die Gruppenzugehörigkeit scheint demnach keinen bzw. höchstens einen marginalen Einfluss auf das Antwortverhalten der Probanden zu haben. (Bühner/Ziegler (2009), S. 335; Eid/Gollwitzer/Schmitt (2015), S. 404)

An der F-verteilt lässt sich ablesen, dass sich zwischen den Versuchsgruppen kein signifikanter Unterschied hinsichtlich des Antwortverhaltens ergibt. Unterschiede zwischen den Gruppen treten somit rein zufällig auf. Die H_0 wird beibehalten und die H_1 abgelehnt.

Durch die Gruppenzugehörigkeit werden nur 0,9% ($\eta^2_{\text{partial}} = .009$) der Varianz in der abhängigen Variable aufgeklärt. Der Effekt liegt mit $f = .095$ knapp unterhalb der Schwelle für einen schwachen Kontrast. (Bühner/Ziegler (2009), S. 366ff; Cohen (1988), S. 284ff; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 194f; Sedlmeier/Renkewitz (2008), S. 452ff)

Die post-hoc-Teststärke liegt mit $1 - \beta = .168$ weit unterhalb des Schwellenwertes von 80%, um das nicht signifikante Ergebnis statistisch absichern zu können. Gleichzeitig ist die β -Fehler-Wahrscheinlichkeit stark erhöht ($\beta > .20$). Die Wahrscheinlichkeit einer Ablehnung der H_0 ist somit sehr gering, während die Wahrscheinlichkeit der Annahme einer gültigen H_1 ebenfalls sehr niedrig ist. Die geringe Stichprobengröße lässt keine eindeutige Interpretation zu. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff, S. 369ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2008), S. 455ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 446ff)

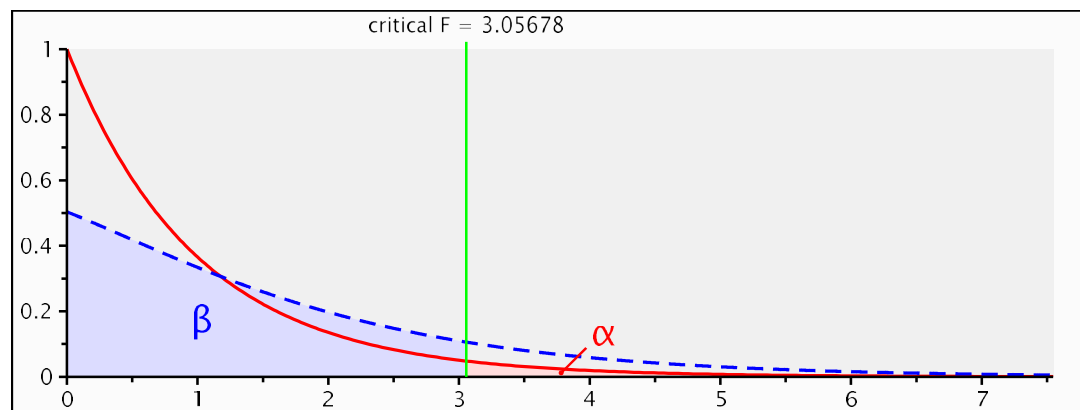


Abb. 7.11: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für $PR_{1(KS)}$ vs. $PT_{2(KS)}$ vs. $PT_{3(KS)}$ bei einseitiger Testung (ANOVA) und $\alpha=5\%$, $df_1=2$, $df_2=149$

7.2.6.4.3 Standard-Skalen: PT_1 vs. PR_2 vs. PT_4

Im Rahmen der Kontrastierung von PT_1 , PR_2 und PT_4 werden ausschließlich Vertrauensmessungen gegenübergestellt, die per Standard-Skala durchgeführt wurden. Dabei kann für PR_2 und PT_4 angenommen werden, dass kein Priming die Vertrauensmessung verzerrt hat. Die Messung in PT_1 kann dagegen durch die Messung in PR_1 beeinflusst sein. Um den Einfluss etwaiger *Pretest*-, *Positions*- oder *Carry-over*-Effekte in PT_1 ausschließen zu können, darf das Antwortverhalten in PT_1 nicht vom Antwortverhalten in PR_2 und PT_4 abweichen. Würde sich ein signifikanter Unterschied ergeben, so müssten Verzerrungen durch die genannten Bias angenommen werden. Es werden die folgenden Hypothesen getestet:

7. Hauptstudie

H₀: Die Gruppen PT₁, PR₂ und PT₄ unterscheiden sich nicht hinsichtlich des Mittelwertes der Standard-Skala zur Erfassung Generalisierten Vertrauens und entstammen daher derselben Population.

$$\mu_{PT1} = \mu_{PR2} = \mu_{PT4}$$

H₁: Mindestens zwei der Gruppen PT₁, PR₂ und PT₄ unterscheiden sich hinsichtlich des Mittelwertes der Standard-Skala zur Erfassung Generalisierten Vertrauens und entstammen daher verschiedenen Populationen.

$$\mu_{PT1} \neq \mu_{PR2} \vee \mu_{PT1} \neq \mu_{PT4} \vee \mu_{PR2} \neq \mu_{PT4}$$

Tab. 7.47: Deskriptive Statistik und Prüfung auf Varianzhomogenität

Gruppe	Deskriptive Statistik			Levene-Test auf Varianzhomogenität	
	\bar{x}	s	N	F	Signifikanz
PT ₁	3,2800	,75701	50	,164	,848 ^{n.s.}
PR ₂	3,3400	,79821	50		
PT ₄	3,1923	,81742	52		
Gesamt	3,2697	,78875	152		

^{n.s.} nicht signifikant (p>.05 (zweiseitig))

(Quelle: eigene Berechnungen)

Anhand des Ergebnisses des Levene-Tests lässt sich ablesen, dass sich die Varianzen innerhalb der Versuchsgruppen nicht unterscheiden und Varianzhomogenität angenommen werden kann. Die gemittelte Gruppenvarianz kann somit als gute Schätzung der Populationsvarianz verwendet werden. Auch die Gruppenmittelwerte scheinen hiernach als erwartungstreue Schätzer für den Populationsmittelwert herangezogen werden können. (Bühner/Ziegler (2009), S. 362f, S. 372ff, S. 378f)

Tab. 7.48: ANOVA (PT₁ vs. PR₂ vs. PT₄)

Gruppe ^d	QS ^a	df ^b	σ^2 ^c
Gruppe ^d	,564	2	,282
Fehler	93,377	149	,627
Total	1719,000	152	---
F ^e		,450	
Signifikanz		,639 ^{n.s.}	
η^2 ^{partial} ^f		,006	
f _g		,078	
1- β		,122 ^h	
β		,878 ⁱ	

^a Quadratsumme

^b Freiheitsgrade („degrees of freedom“): $df_{treat} = k - 1$; $df_{fehler} = k * (n - 1)$; $df_{tot} = n * k - 1$ mit k=Anzahl der Faktorstufen; n=Fallzahl (Bortz (2005), S. 251ff)

^c Mittel der Quadrate

^d Treatment

^e F-verteilte Prüfgröße

^f Effektstärke bzw. Anteil aufgeklärter Varianz an der Gesamtvarianz: $\eta^2_{partial} = \frac{QS_{treat}}{QS_{treat} + QS_{fehler}}$; $\eta^2 \approx .01 \rightarrow$ „kleiner“ Effekt, $\eta^2 \approx .06 \rightarrow$ „mittlerer“ Effekt, $\eta^2 \approx .14 \rightarrow$ „großer“ Effekt (Eid/Gollwitzer/Schmitt (2015), S. 409f; Bortz (2005), S. 255, S. 259, S. 280; Bühner/Ziegler (2009), S. 366ff)

^g Effektstärke: $f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$; $f = .10 \rightarrow$ kleiner Effekt; $f = .25 \rightarrow$ moderater Effekt; $f = .40 \rightarrow$ großer Effekt (Bühner/Ziegler (2009),

S. 368; Cohen (1988), S. 284ff; Eid/Gollwitzer/Schmitt (2015), S. 411f)

^h Teststärke bzw. beobachtete Schärfe für $\alpha = 5\%$, $df_1 = 2$ u. $df_2 = 149$ (Bühner/Ziegler (2009), S. 370, S. 381)

ⁱ β -Fehler-Wahrscheinlichkeit für $\alpha = 5\%$, $df_1 = 2$; $df_2 = 149$: „Die Irrtumswahrscheinlichkeit β ist die Wahrscheinlichkeit, mit der ein statistischer Test ein nicht-signifikantes Ergebnis ergibt, obwohl in Wirklichkeit die H₁ gilt.“ (Eid/Gollwitzer/Schmitt (2015), S. 224)

^{n.s.} nicht signifikant (p>.05 (zweiseitig))

(Quelle: eigene Berechnungen)

Stellt man die $Q_{S_{treat}}$ und $Q_{S_{fehler}}$ gegenüber wird sichtbar, dass Abweichungen vom Mittelwert der Grundgesamtheit vorwiegend aus Variationen innerhalb der Versuchsgruppen resultieren, allerdings kaum Unterschiede zwischen den Gruppen bestehen. Die Differenzen innerhalb der Gruppen tragen deutlich mehr zur Gesamtvariation bei, als dies die Abweichungen aufgrund von Unterschieden zwischen den Gruppen tun. (Bühner/Ziegler (2009), S. 342f; Eid/Gollwitzer/Schmitt (2015), S. 396f; Sedlmeier/Renkewitz (2008), S. 443ff; Sedlmeier/Renkewitz (2013), S. 437f)

Außerdem lässt sich feststellen, dass die Varianz innerhalb der Gruppen deutlich höher ausfällt als die Varianz zwischen den Gruppen. Die Gruppenzugehörigkeit scheint somit keinen Effekt auf das Antwortverhalten der Versuchspersonen zu haben. (Bühner/Ziegler (2009), S. 335; Eid/Gollwitzer/Schmitt (2015), S. 404)

Als Bestätigung der bisherigen Beobachtungen zeigt die F-verteilte Prüfgröße keinen signifikanten Unterschied im Antwortverhalten zwischen den Gruppen an. Es sollte daher die H_0 beibehalten und die H_1 abgelehnt werden.

Durch die Gruppenmitgliedschaft werden lediglich 0,6% ($\eta^2_{\text{partial}} = .006$) der Varianz innerhalb der abhängigen Variable „Generalisiertes Vertrauen“ aufgeklärt. Die Effektstärke zeigt mit $f = .078$ einen Kontrast unterhalb der Schwelle für einen kleinen Effekt an. Die Differenz im Antwortverhalten zwischen den Versuchsgruppen kann daher als marginal bewertet werden. (Bühner/Ziegler (2009), S. 366ff; Cohen (1988), S. 284ff; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 194f; Sedlmeier/Renkewitz (2008), S. 452ff)

Für die post-hoc-Teststärke ergibt sich bei $\alpha = 5\%$ ein Wert von $1 - \beta = .122$ und eine β -Fehler-Wahrscheinlichkeit von $\beta = .878$. Die Teststärke fällt damit zu gering aus, als dass eine statistische Absicherung des nicht signifikanten Ergebnisses möglich wäre. Die Wahrscheinlichkeit, eine gültige H_1 beibehalten zu können, ist zu gering. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff, S. 369ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2008), S. 455ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 446ff)

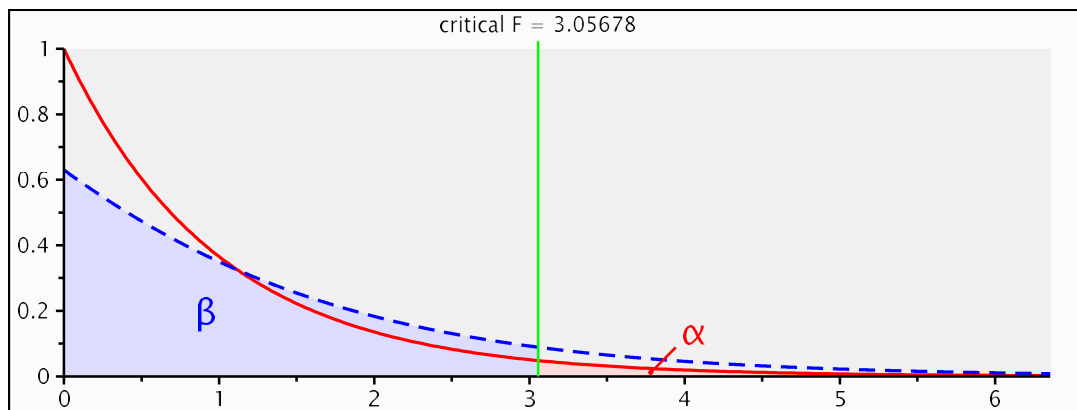


Abb. 7.12: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PT_1 vs. PR_2 vs. PT_4 bei einseitiger Testung (ANOVA) und $\alpha = 5\%$, $df_1 = 2$, $df_2 = 149$

7.2.6.5 Nonparametrische Signifikanzprüfung

Im Einklang mit der Parallelprüfungsempfehlung von BORTZ/LIENERT/BOEHNKE sollen im nachfolgenden Abschnitt die Ergebnisse der vorangegangenen Signifikanzprüfung mit Hilfe parametrischer Verfahren anhand von verteilungsfreien Verfahren validiert werden. Da es sich

7. Hauptstudie

bei den in die Signifikanzprüfung eingehenden Versuchsgruppen um wechselseitig unabhängige Stichproben handelt, kommt zur Ermittlung der Unterschiede zwischen den Objektiv-Skalen, den Objektiv-Kurzskalen und den Standard-Skalen der Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben zum Einsatz. Es ergeben sich dann die folgenden Hypothesen:

H_0 : Die Mediane der Objektiv-Skalen unterscheiden sich wechselseitig nicht voneinander bzw. die Mediane der Objektiv-Kurzskalen unterscheiden sich wechselseitig nicht voneinander bzw. die Mediane der Standard-Skalen unterscheiden sich wechselseitig nicht voneinander. In der Grundgesamtheit liegen jeweils gleich viele Rangplatzüberschreitungen wie –unterschreitungen vor.

$$\tilde{x}_{PR} = \tilde{x}_{PT2} \text{ bzw. } \tilde{x}_{PR1(KS)} = \tilde{x}_{PT2(KS)}$$

$$\tilde{x}_{PT} = \tilde{x}_{PT} \text{ bzw. } \tilde{x}_{PT2(KS)} = \tilde{x}_{PT3(KS)}$$

$$\tilde{x}_{PT} = \tilde{x}_{PR2}$$

$$\tilde{x}_{PT1} = \tilde{x}_{PT4}$$

H_1 : Die Mediane der Objektiv-Skalen unterscheiden sich wechselseitig voneinander bzw. die Mediane der Objektiv-Kurzskalen unterscheiden sich wechselseitig voneinander bzw. die Mediane der Standard-Skalen unterscheiden sich wechselseitig voneinander. In der Grundgesamtheit liegen jeweils ungleich viele Rangplatzunterschreitungen und -überschreitungen vor.

$$\tilde{x}_{PR1} \neq \tilde{x}_{PT2} \text{ bzw. } \tilde{x}_{PR1(KS)} \neq \tilde{x}_{PT2(KS)}$$

$$\tilde{x}_{PT2} \neq \tilde{x}_{PT3} \text{ bzw. } \tilde{x}_{PT(KS)} \neq \tilde{x}_{PT3(KS)}$$

$$\tilde{x}_{PT1} \neq \tilde{x}_{PR2}$$

$$\tilde{x}_{PT1} \neq \tilde{x}_{PT}$$

Die Kontrastierung der Objektiv-Skalen bzw. Objektiv-Kurzskalen bzw. Standard-Skalen ergibt jeweils keinen überzufälligen Unterschied. Allerdings können die Ergebnisse nicht statistisch abgesichert werden. Die geringe Teststärke geht dabei mit einer β -Fehler-Wahrscheinlichkeit weit oberhalb des akzeptierten β -Fehler-Niveaus von $\beta=.20$ einher. Es kann daher keine eindeutige Interpretation des Ergebnisses geleistet werden. Die Datengrundlage des Signifikanztests scheint zu gering zu sein, als dass eine Entscheidung zu Gunsten der Alternativhypothese getroffen werden konnte. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärken liegen, mit Ausnahme des Effektes von $PR_{1(KS)}$ vs. $PT_{2(KS)}$, unterhalb der Schwelle für schwache Effekte ($\omega = .10$). Zwischen den Messergebnissen der Skalen scheinen demnach nur geringe Differenzen zu bestehen bzw. die Differenzen sind in der Praxis als irrelevant zu bewerten. Dies spricht dafür, dass das nicht-signifikante Ergebnis unabhängig von der niedrigen Teststärke Gültigkeit besitzt. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 668f; Bühner/Ziegler (2009), S. 177ff, S. 269f, S. 633; Cohen (1988), S. 9f, S. 223ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Die mittleren Ränge der verschiedenen Skalen unterscheiden sich jeweils nur marginal. (Bühner/Ziegler (2009), S. 287)

Entsprechend der z-verteilten Prüfgrößen wird für jede der Signifikanzprüfungen die H_0 beibehalten und die H_1 abgelehnt. Auf den Objektiv-Skalen bzw. Objektiv-Kurzskalen und

7. Hauptstudie

Standard-Skalen resultiert somit unabhängig von der Versuchsgruppe und des Expositionszeitpunktes jeweils ein ähnliches bzw. das gleiche Messergebnis. Allerdings muss berücksichtigt werden, dass aufgrund der geringen Teststärke und der hohen β -Fehler-Wahrscheinlichkeit das Risiko eines β -Fehlers nicht ausgeschlossen ist.

Tab. 7.49: Nonparametrische Verfahren

Pretest-, Positions- und Carry-over-Effekt	
<i>Objektiv-Skala</i>	
PR ₁ vs. PT ₂	$z = -.799$ (zweiseitig); n.s.; $\varphi = .080$; $1-\beta = .067$
PT ₂ vs. PT ₃	$z = -.533$ (zweiseitig); n.s.; $\varphi = .053$; $1-\beta = .058$
<i>Objektiv-Kurzskala</i>	
PR _{1(KS)} vs. PT _{2(KS)}	$z = -1.143$ (zweiseitig); n.s.; $\varphi = .114$; $1-\beta = .085$
PT _{2(KS)} vs. PT _{3(KS)}	$z = -.532$ (zweiseitig); n.s.; $\varphi = .053$; $1-\beta = .058$
<i>Standard-Skala</i>	
PT ₁ vs. PR ₂	$z = -.264$ (zweiseitig); n.s.; $\varphi = .026$; $1-\beta = .052$
PT ₁ vs. PT ₄	$z = -.509$ (zweiseitig); n.s.; $\varphi = .050$; $1-\beta = .057$

(Quelle: Eigene Berechnungen)

7.2.6.6 Zusammenfassung

Im Rahmen der Prüfung auf potentielle Messverzerrungen durch Effekte der experimentellen Situation wurde geprüft, ob sich die einzelnen Objektiv-Messungen signifikant voneinander unterscheiden bzw. ob sich die einzelnen Standard-Messungen signifikant voneinander unterscheiden. Messungen, denen eine Befragung vorangegangen war, unabhängig davon ob es sich um die Objektiv- oder die Standard-Skala handelte, mussten dabei als mit einem Bias belastet angenommen werden. Dagegen konnten Messungen, denen keine Befragung vorangegangen war, als *one-shot-case-study* interpretiert werden und daher Biasfreiheit angenommen werden. Als potentiell mit einem Bias belastet wurde daher auf Seiten der Objektiv-Skala die Messung in PT₂ bzw. PT_{2(KS)} eingestuft, während auf Seiten der Standard-Skala die Messung in PT₁ als potentiell mit einem Bias belastet gewertet wurde. Als von einem Bias frei können dagegen PR₁ bzw. PR_{1(KS)} und PT₃ bzw. PT_{3(KS)} für die Objektiv-Skala und PR₂ und PT₄ für die Standard-Skala angesehen werden.

Damit eine Beeinflussung des Unterschiedes zwischen Objektiv- und Standard-Skala zur Erfassung Generalisierten Vertrauens durch etwaige *Pretest*-, *Positions*- und *Carry-over*-Effekte ausgeschlossen werden konnte, durfte sich das Antwortverhalten bei Messungen, denen bereits eine Befragung vorausgegangen war, nicht vom Antwortverhalten bei Messungen, denen keine Befragung vorangegangen war, unterscheiden. Ein signifikanter Unterschied im Antwortverhalten würde dagegen anzeigen, dass die jeweiligen Vorher-Messungen einen Einfluss auf die Probanden ausgeübt haben.

Betrachtet man das Ergebnis der parametrischen Analyse (hier: einfaktorielle Varianzanalyse), so ergibt sich weder für die Objektiv-Skala, die Objektiv-Kurzskala noch die Standard-Skala ein signifikanter Unterschied zwischen den Versuchsgruppen. Die Effektstärke nimmt Werte von $.095 \leq f \leq .105$ an, was einem kleinen Effekt entspricht. Der Kontrast zwischen den Gruppen ist entsprechend schwach ausgeprägt. Ausgehend von der post-hoc-Teststärke, die Werte von $.168 \leq 1-\beta \leq .192$ annimmt, kann die Übereinstimmung des Antwortverhaltens über alle Versuchsgruppen hinweg allerdings zunächst nicht statistisch abgesichert werden. Das hierfür erforderliche Teststärkeniveau von $1-\beta = .80$ wird stark unterschritten bzw. die Wahrscheinlichkeit zugunsten der Alternativhypothese (H_1) zu entscheiden, wenn diese in der

7. Hauptstudie

Grundgesamtheit Gültigkeit besitzt, ist zu gering. Voraussichtlich ist die in die Analyse eingehende Fallzahl zu gering, um eine eindeutige Interpretation des Testresultates zu erlauben.

Tab. 7.50: Parametrische Verfahren (ANOVA)

Pretest-, Positions- und Carry-over-Effekt	
PR₁ vs. PT₂ vs. PT₃	F(2;149)=,838; n.s.; f=,105; 1-β=,192
PR_{1(KS)} vs. PT_{2(KS)} vs. PT_{3(KS)}	F(2;149)=,838; n.s.; f=,105; 1-β=,192
PT₁ vs. PR₂ vs. PT₄	F(2;149)=,708; n.s.; f=,095; 1-β=,168

F(df₁;df₂): empirischer F-Wert inklusive Zähler- und Nennerfreiheitsgrade

n.s.: nicht signifikant (p>.05 (zweiseitig))

f: Effektstärke $f = .10 \rightarrow$ kleiner Effekt; $f = .25 \rightarrow$ moderater Effekt; $f = .40 \rightarrow$ großer Effekt

1-β: Teststärke

Vgl. Eid/Gollwitzer/Schmitt (2015), S. 407

(Quelle: eigene Berechnungen)

Um trotz der geringen Fallzahl eine belastbare Aussage hinsichtlich der Beeinflussung der Messungen durch Effekte der experimentellen Situation machen zu können, wurde zusätzlich anhand verteilungsfreier Verfahren geprüft, ob sich das Antwortverhalten auf den Objektiv- bzw. Standard-Skalen über die Gruppen hinweg unterscheidet.

Auch die nonparametrische Analyse (hier: Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben) kommt zu dem Ergebnis, dass sich weder für die Objektiv-Skala, die Objektiv-Kurzskaala noch die Standard-Skala ein signifikanter Unterschied zwischen den Versuchsgruppen ergibt. Die Effektstärke bewegt sich mit $.026 \leq \varphi \leq .114$ auf einem ähnlichen Niveau wie bereits bei der parametrischen Analyse. Der Kontrast zwischen den Versuchsgruppen kann somit als marginal bezeichnet werden. Die post-hoc-Teststärke rangiert mit $.083 \leq 1-\beta \leq .307$ weit unterhalb des notwendigen Niveaus von $1-\beta=.80$. Für sich genommen kann somit auch anhand der nonparametrischen Verfahren das Resultat der Signifikanzprüfung nicht statistisch abgesichert werden.

Die Überprüfung durch die nonparametrischen Verfahren bestätigt das parametrische Testergebnis, was die endgültige Beibehaltung der Nullhypothese (H_0) zur Folge hat. (Bortz/Lienert/Boehnke (2008), S. 85f) Es wird demnach davon ausgegangen, dass hinsichtlich des Antwortverhaltens auf den unterschiedlichen Vertrauensskalen kein signifikanter Unterschied zwischen den Versuchsgruppen besteht. Daraus lässt sich schließen, dass die Resultate der Messungen in PT₂, PT_{2(KS)} und PT₁ nicht von den Resultaten ihrer jeweiligen als mit einem Bias belastet angenommenen Pendants abweichen. Außerdem kann nun davon ausgegangen werden, dass innerhalb des Surveyexperimentes keine Beeinflussung der verschiedenen Messungen durch Pretest-, Positions- oder Carry-over-Effekte stattgefunden hat. Dies stellt gleichzeitig einen Hinweis darauf dar, dass die im Rahmen der Objektiv-Messung realisierte geringe Augenscheinvalidität hinsichtlich der Zieldimension „Generalisiertes Vertrauen“ wirksam war.

7.2.6.7 Effekt des Treatments (Gesamt N)

Im folgenden Abschnitt wird geprüft, ob sich das per Objektiv-Skalen abgebildete Vertrauensniveau von dem per Standard-Skala ermittelten Vertrauensniveau unterscheidet. Dabei gehen alle gültigen Fälle in die Analyse ein. Außerdem wird nach Within-subject- und Between-subject-Design unterschieden und es werden sowohl parametrische als auch nonparametrische Signifikanzprüfungen durchgeführt.

7.2.6.7.1 Within-subject-Design

7.2.6.7.1.1 Gruppe 1: Objektiv/Standard (PR₁ vs. PT₁)

Die Probanden der ersten Versuchsgruppe bearbeiteten sowohl die Objektiv- als auch die Standard-Skala. Wobei zunächst die Objektiv-Skala präsentiert wurde, um Priming-Effekte zu vermeiden, welche die im Zuge der Objektiv-Messung realisierte geringe Augenscheinvalidität bezüglich der Zieldimension „Generalisiertes Vertrauen“, unterwandern könnten. Unter der Bedingung, dass der Posttest in der ersten Versuchsgruppe frei von Einflüssen durch den vorangegangenen Pretest ist, lässt sich dann durch die Gegenüberstellung von PR₁ und PT₁ der Treatment-Effekt isolieren.

7.2.6.7.1.1.1 Parametrische Signifikanzprüfung

Mit Hilfe des T-Tests für abhängige Stichproben soll nun überprüft werden, ob sich hinsichtlich des Antwortverhaltens der Probanden auf den verschiedenen Skalen zur Erfassung Generalisierten Vertrauens signifikante Unterschiede ergeben bzw. ob die Probanden auf der Objektiv-Skala einen signifikant niedrigeren Vertrauenswert erreichen als auf der Standard-Skala. Dabei entspricht die Objektiv-Variante Messzeitpunkt 1 (PR₁) und die Standard-Variante entspricht Messzeitpunkt 2 (PT₁). Es werden dementsprechend die folgenden Hypothesen getestet:

H₀: Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR₁) zur Erfassung Generalisierten Vertrauens und der Standard-Skala (PT₁) zur Erfassung Generalisierten Vertrauens fallen innerhalb der Population gleich aus bzw. die Differenz der Messwertpaare in der Grundgesamtheit ist gleich null. Abweichungen zwischen den Mittelwerten von Objektiv-Skala bzw. -Kurzskala und Standard-Skala sind rein zufällig.

$$\mu_{PR1} = \mu_{PT1} \text{ bzw. } \mu_d = 0$$

H₁: Der Mittelwert der Objektiv-Skala bzw. -Kurzskala (PR₁) zur Erfassung Generalisierten Vertrauens fällt innerhalb der Population jeweils niedriger aus als der Mittelwert der Standard-Skala (PT₁) zur Erfassung Generalisierten Vertrauens bzw. die Differenz der Messwertpaare innerhalb der Grundgesamtheit ist kleiner null.

$$\mu_{PR1} < \mu_{PT1} \text{ oder } \mu_d < 0$$

Tab. 7.51: T-Test für abhängige Stichproben (PR₁ vs. PT₁)

	N	\bar{x}	s
Objektiv (PR₁)	50	2,880	,52060
Standard (PT₁)	50	3,280	,75701
\bar{x}_d^a	-,400		
$\hat{\sigma}_d^b$,60609		
Konf_{95%} (Differenz)^c	-,57225		-,22775
T^d	-4,667		
Df^e	49		
Signifikanz	,000**		
Hedges g/Cohens d^f	,660		
1-β	,998 ^g		,986 ^h

^a Mittelwert der Differenzen^b Standardabweichung der Differenzen^c 95% Konfidenzintervall der Differenz; Links: Untere Grenze, Rechts: Obere Grenze^d t-verteilte Prüfgröße^e Freiheitsgrade (*degrees of freedom*) mit $df=n-1$ (Eid/Gollwitzer/Schmitt (2015), S. 370; Bortz (2005), S. 144)^f Effektstärke bzw. Effektgröße: Hedges g bzw. Cohens d_s, mit g bzw. $d_s = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}_d} = \frac{|\bar{x}_d|}{\hat{\sigma}_d}$, $d = .20 \rightarrow$ *kleiner Effekt*, $d = .50 \rightarrow$ *mittlerer Effekt*, $d = .80 \rightarrow$ *großer Effekt* (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)^g post-hoc-Teststärke (*power*) für $\alpha=5\%$, $df=49$: Wahrscheinlichkeit, mit der zu Gunsten der H₁ entschieden wird, wenn ein Unterschied besteht bzw. die H₀ gilt (Bortz (2005), S. 123; Bühner/Ziegler (2009), S. 247, S. 252)^h post-hoc-Teststärke („*power*“) für $\alpha=1\%$, $df=49$ * $p<.05$ (einseitig), ** $p<.01$ (einseitig)

(Quelle: eigene Berechnungen)

Ein Augenscheinvergleich der Mittelwerte der verschiedenen Vertrauensmessung zeigt bereits, dass die Probanden auf der Objektiv-Skala ein geringeres Generalisiertes Vertrauen angeben als auf der Standard-Skala. Ausgehend hiervon scheint ein Unterschied im Antwortverhalten je nach Messvariante zu bestehen, der von der Alternativhypothese postulierten Richtung entspricht.

Die t-verteilte Prüfgröße bestätigt diese Beobachtung und zeigt ein hochsignifikantes Ergebnis an. Der Unterschied zwischen den beiden Messverfahren zur Erfassung Generalisierten Vertrauens ist damit als überzufällig zu bewerten. Es sollte demnach die H₀ abgelehnt und die H₁ beibehalten werden.

Die Effektstärke kann mit $g/d_s=.660$ als mittlerer Effekt ($d \geq .50$) eingeordnet werden. Es besteht somit eine praktisch bedeutsame Differenz zwischen Objektiv- und Standard-Messung, die auch in der Grundgesamtheit Bestand hat. Diese Interpretation wird zusätzlich durch das 95%-Konfidenzintervall der Differenz gestützt. So bewegen sich die Werte der mittleren Differenz mit 95% Wahrscheinlichkeit im Bereich von $-.572 < \bar{x}_d < -.228$. Der Wert null ist in diesem Konfidenzintervall nicht enthalten. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 187, S. 245, S. 250; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 373f)

Für die post-hoc-Teststärke ergibt sich auf dem 5%-Signifikanzniveau ein Wert von $1-\beta=.998$, während auf dem 1%-Signifikanzniveau ein ähnlicher Wert von $1-\beta=.986$ erreicht wird. Das signifikante Ergebnis kann somit auf beiden Signifikanzniveaus als statistisch abgesichert angesehen werden. Die Wahrscheinlichkeit die H₁ irrtümlich abzulehnen, ist als sehr gering zu bewerten. Mit anderen Worten: Die Wahrscheinlichkeit zu Gunsten der H₁ zu entscheiden, wenn tatsächlich ein Unterschied besteht, ist sehr hoch bzw. oberhalb des notwendigen Schwellenwertes von $1-\beta=.80$. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

7. Hauptstudie

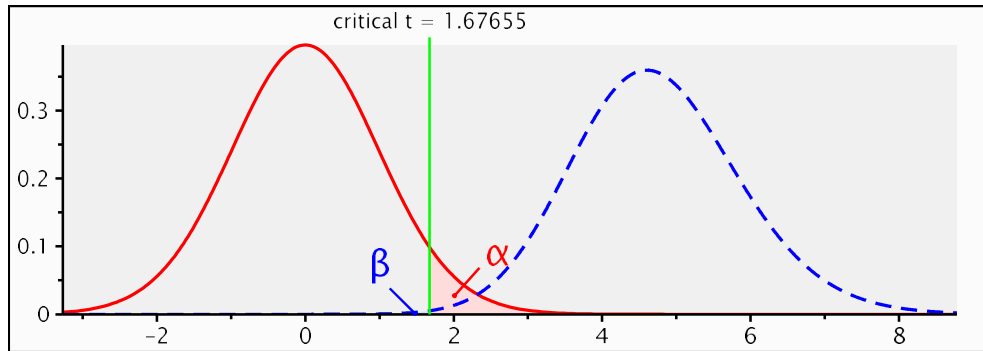


Abb. 7.13: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PR_1 (rot) vs. PT_1 (blau) bei einseitiger Testung (T-Test für abhängige Stichproben) und $\alpha=5\%$, $df=49$

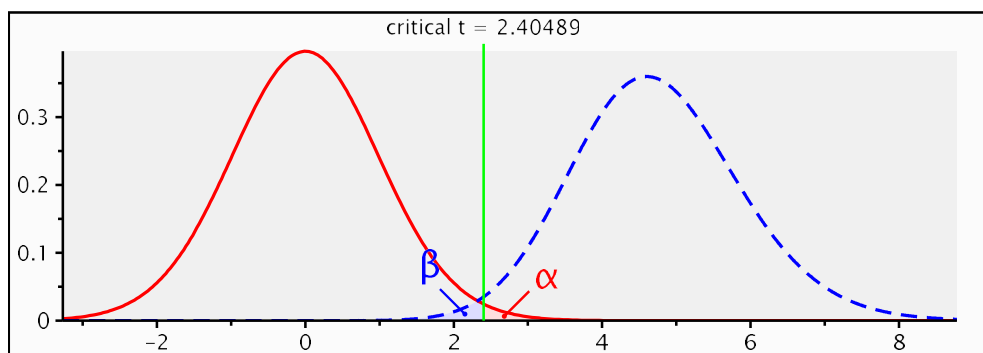


Abb. 7.14: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für $PR_{1(KS)}$ (rot) vs. PT_1 (blau) bei einseitiger Testung (T-Test für abhängige Stichproben) und $\alpha=1\%$, $df=49$

Im nächsten Schritt soll überprüft werden, ob auch für das Verhältnis von Objektiv-Kurzskala und Standard-Skala die gleiche Beziehung wie für Objektiv- und Standard-Skala gilt.

Tab. 7.52: T-Test für abhängige Stichproben ($PR_{1(KS)}$ vs. PT_1)

	N	\bar{x}	s
Objektiv ($PR_{1(KS)}$)	50	3,000	,63888
Standard (PT_1)	50	3,280	,75701
\bar{x}_d^a		-,28000	
$\hat{\sigma}_d^b$,57286	
Konf _{95%} (Differenz) ^c		-,44280	-,11720
T ^d		-3,456	
Df ^e		49	
Signifikanz		,001**	
Hedges g/Cohens d ^f		,489	
1- β		,961 ^g	,850 ^h

^a Mittelwert der Differenzen

^b Standardabweichung der Differenzen

^c 95% Konfidenzintervall der Differenz; Links: Untere Grenze, Rechts: Obere Grenze

^d t-verteilte Prüfgröße

^e Freiheitsgrade (*degrees of freedom*) mit $df=n-1$ (Eid/Gollwitzer/Schmitt (2015), S. 370; Bortz (2005), S. 144)

^f Effektstärke bzw. Effektgröße: Hedges g bzw. Cohens d_s, mit g bzw. $d_s = \frac{|x_1 - x_2|}{\hat{\sigma}_d} = \frac{|x_d|}{\hat{\sigma}_d}$, $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

^g post-hoc-Teststärke (*power*) für $\alpha=5\%$, $df=49$: Wahrscheinlichkeit, mit der zu Gunsten der H_1 entschieden wird, wenn ein Unterschied besteht bzw. die H_1 gilt (Bortz (2005), S. 123; Bühner/Ziegler (2009), S. 247, S. 252)

^h post-hoc-Teststärke („power“) für $\alpha=1\%$, $df=49$

* $p<.05$ (einseitig), ** $p<.01$ (einseitig)

(Quelle: eigene Berechnungen)

7. Hauptstudie

Betrachtet man die Mittelwerte der beiden Messvarianten, so erreichen die Probanden auf der Objektiv-Kurzskala ein geringeres Generalisiertes Vertrauen als auf der Standard-Skala. Die Richtung des potentiellen Unterschiedes zwischen den beiden Messvarianten entspricht demnach der von der Alternativhypothese vorhergesagten Tendenz.

An der t-verteilten Prüfgröße lässt sich ein hochsignifikanter Unterschied zwischen den Messmethoden ablesen. Es sollte demnach die H_0 abgelehnt und die H_1 beibehalten werden.

Für die Effektstärke ergibt sich ein Wert von $g/d_s = .489$, was einem kleinen bis mittleren Effekt entspricht. Es existiert somit eine relevante Differenz zwischen den Mittelwerten der beiden Messvarianten. Das 95%-Konfidenzintervall beinhaltet außerdem nicht den Wert null, was einen weiteren Beleg für den signifikanten Kontrast zwischen den beiden Messmethoden darstellt. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 187, S. 245, S. 250; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 373f)

Die post-hoc-Teststärke errechnet sich sowohl für $\alpha=5\%$ als auch für $\alpha=1\%$ zu einem Wert oberhalb des notwendigen Teststärkeniveaus von $1-\beta=.80$. Das signifikante Ergebnis des T-Tests kann somit als statistisch abgesichert angesehen werden. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

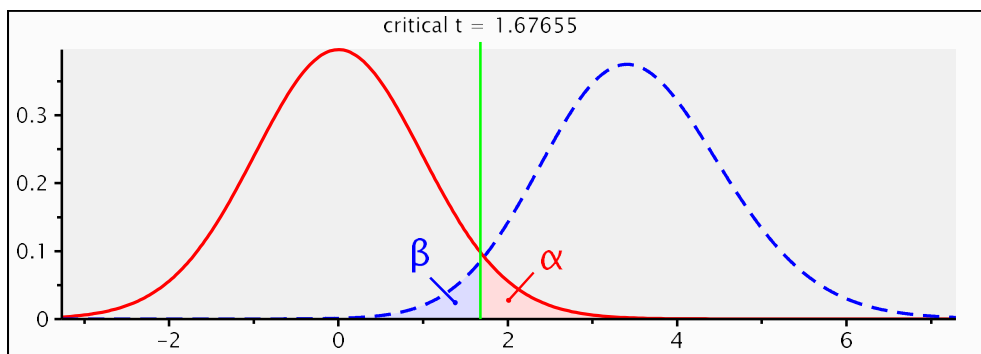


Abb. 7.15: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für $PR_{1(KS)}$ (rot) vs. PT_1 (blau) bei einseitiger Testung (T-Test für abhängige Stichproben) und $\alpha=5\%$, $df=49$

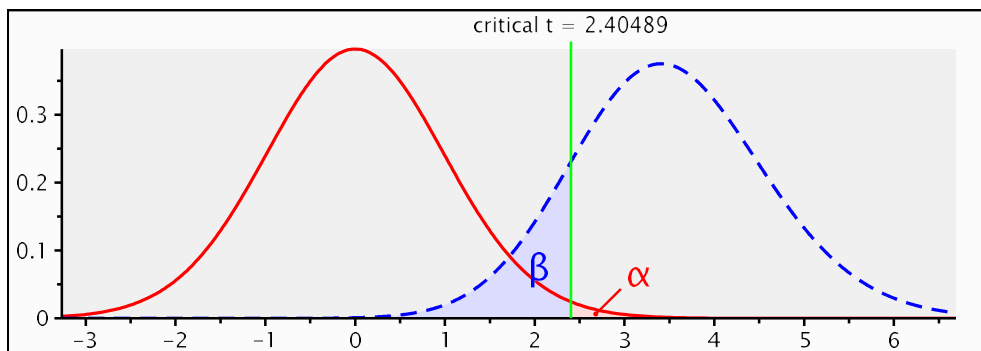


Abb. 7.16: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für $PR_{1(KS)}$ (rot) vs. PT_1 (blau) bei einseitiger Testung (T-Test für abhängige Stichproben) und $\alpha=1\%$, $df=49$

7.2.6.7.1.1.2 Nonparametrische Signifikanzprüfung

Innerhalb der ersten Versuchsgruppe bearbeiteten die Probanden zunächst die Objektiv-Skala und danach die Standard-Skala. Diese Reihenfolge der Präsentation wurde gewählt, um Priming-

7. Hauptstudie

Effekte bzw. vorwärts gerichtete Positionseffekte, welche die für die Objektiv-Messung angestrebte niedrige Augenscheinvalidität bezüglich der Zieldimension „Generalisiertes Vertrauen“ unterminieren könnten, auszuschließen. So sollte auf diese Weise verhindert werden, dass die Messintention der Objektiv-Skala aufgedeckt würde und die Probanden daraufhin auf beiden Skalen konsistent antworten würden. Durch die gewählte Bearbeitungsreihenfolge kann allerdings nicht ausgeschlossen werden, dass die Probanden ihr Antwortverhalten auf der Standard-Skala an ihr Antwortverhalten auf der Objektiv-Skala anpassen. Eine Kontrolle dieses Effektes erfolgt anhand der zweiten Versuchsgruppe, innerhalb derer die umgekehrte Bearbeitungsreihenfolge realisiert wurde.

Innerhalb der ersten Versuchsgruppe soll nun mit Hilfe des Vorzeichen-Rang-Tests von WILCOXON für abhängige Stichproben überprüft werden, ob sich das Antwortverhalten der Probanden zwischen den beiden Messvarianten zur Erfassung Generalisierten Vertrauens signifikant unterscheidet bzw. ob die Probanden auf der Objektiv-Skala einen signifikant niedrigeren Vertrauenswert erreichen als auf der Standard-Skala. Entlang der Bearbeitungsreihenfolge entspricht die Objektiv-Variante Messzeitpunkt 1 (PR₁) und die Standard-Variante entspricht Messzeitpunkt 2 (PT₁). Es werden dementsprechend die folgenden Hypothesen getestet:

H₀: Die Mediane der Objektiv-Skala bzw. -Kurzskala (PR₁) zur Erfassung Generalisierten Vertrauens und der Standard-Skala (PT₁) zur Erfassung Generalisierten Vertrauens fallen in der Grundgesamtheit gleich aus bzw. die Rangsummen der beiden Messvarianten sind in der Population gleich. Unterschiede zwischen der Objektiv-Skala bzw. -Kurzskala und der Standard-Skala sind rein zufällig.

$$W^+ = W^- \text{ bzw. } \tilde{x}_{PR1} = \tilde{x}_{PT1}$$

H₁: Der Median der Objektiv-Skala bzw. -Kurzskala (PR₁) fällt innerhalb der Grundgesamtheit niedriger aus, als der Median der Standard-Messung (PT₁). Zwischen den beiden Messvarianten treten positive Ränge mit größerer Häufigkeit auf als negative Ränge.

$$W^+ > W^- \text{ bzw. } \tilde{x}_{PR1} < \tilde{x}_{PT1}$$

Tab. 7.53: Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben (PR₁ vs. PT₁)

	N	Mittlerer Rang	Rangsumme
Negative Ränge ^a	3	13,50	40,50 ^d
Positive Ränge ^b	23	13,50	310,50 ^e
Bindungen ^c	24	---	---
Gesamt	50	---	---
Z ^f	-3,922		
Exakte Signifikanz	,000**		
φ ^g	,555		
1- β	,984 ^h		,919 ⁱ

^a Standard < Objektiv

^b Standard > Objektiv

^c Standard = Objektiv

^d W⁻

^e W⁺

^f z-verteilte Prüfgröße, basierend auf negativen Rängen

^g Effektstärke bzw. Effektgröße: $\varphi = \sqrt{\frac{z^2}{N}} = \frac{z}{\sqrt{N}}$ mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

^h post-hoc-Teststärke („power“) für $\alpha=5\%$, $df=46,7464829$: Wahrscheinlichkeit, mit der zu Gunsten der H₁ entschieden wird, wenn ein Unterschied besteht bzw. die H₁ gilt (Bortz (2005), S. 123; G* Power 3.1 manual (2007), S. 52ff)

ⁱ post-hoc-Teststärke („power“) für $\alpha=1\%$, $df=46,7464829$

*p<.05 (einseitig), **p<.01 (einseitig)

(Quelle: eigene Berechnungen)

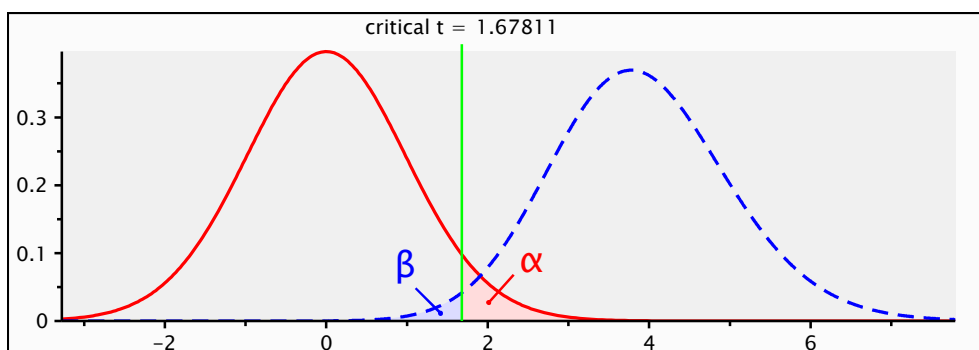
Zunächst ist festzuhalten, dass der Median der Standard-Messung wie auch der Objektiv-Messung jeweils den Wert 3 annimmt. Ausgehend von der bloßen Betrachtung der Mediane kann somit kein Unterschied zwischen den beiden Messvarianten festgestellt werden.

Das Ergebnis des Wilcoxon-Tests zeigt allerdings, dass in 3 Fällen der mit Hilfe der Standard-Variante ermittelte Vertrauenswert niedriger ist als der mit Hilfe der Objektiv-Variante ermittelte Vertrauenswert. Gleichzeitig ergibt sich für 23 Fälle die umgekehrte Relation. Die verbleibenden 24 Fälle unterscheiden sich im angegebenen Vertrauen nicht zwischen der Standard- und der Objektiv-Messung. Entsprechend ergibt sich für positive Ränge eine höhere Rangsumme als für negative Ränge bzw. es sprechen mehr Paardifferenzen für eine Beibehaltung der H_1 als für eine Beibehaltung der H_0 .

Die auf Basis der negativen Ränge ermittelte z-verteilte Prüfgröße bestätigt diesen Eindruck und zeigt ein hochsignifikantes Ergebnis an. Der Unterschied zwischen der Standard-Vertrauensmessung und der Objektiv-Vertrauensmessung ist damit als überzufällig zu bewerten und entspricht der in der Alternativhypothese postulierten Richtung. Es sollte demnach die H_1 beibehalten und die H_0 abgelehnt werden.

Zur statistischen Absicherung bzw. zur Bestimmung der praktischen Bedeutsamkeit des Ergebnisses des Wilcoxon-Tests wurden die Effektstärke und die post-hoc-Teststärke bestimmt. Die Effektstärke erreicht einen Wert von $\varphi = .555$, was einem starken Effekt entspricht ($\omega \geq .50$). Der Unterschied zwischen Standard- und Objektiv-Messung kann somit als praktisch bedeutsam interpretiert werden. Mit anderen Worten: Bei der Differenz der Messvarianten handelt es sich um einen substanziellen Effekt bzw. der Abstand zwischen dem per Standard-Variante ermittelten Vertrauenswert und dem per Objektiv-Variante ermittelten Vertrauenswert ist auch in der Population existent. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 270, S. 633; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Unter der Bedingung, dass der empirische z-Wert ($z=-3,922$) als Erwartungswert der Alternativhypothesenverteilung angenommen wird und der Annahme eines 5%-Signifikanzniveaus, errechnet sich die post-hoc-Teststärke zu $1-\beta = .984$. Die Wahrscheinlichkeit die H_1 irrtümlich abzulehnen, ist damit als sehr gering zu bewerten. (Bühner/Ziegler (2009), S. 188ff; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f) Anders ausgedrückt ist die Wahrscheinlichkeit zu Gunsten der H_1 zu entscheiden, wenn tatsächlich ein Unterschied besteht sehr hoch bzw. oberhalb des notwendigen Schwellenwertes von $1-\beta=.80$. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Cohen (1988), S. 4; Cohen (1992), S. 156) Auch eine konservativere Testung mit $\alpha=1\%$ verändert die post-hoc-Teststärke nur marginal ($1-\beta=0,919$).



7. Hauptstudie

Abb. 7.17: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PR_1 (rot) vs. PT_1 (blau) bei einseitiger Testung (Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben) und $\alpha=5\%$, $df=46,7464829$

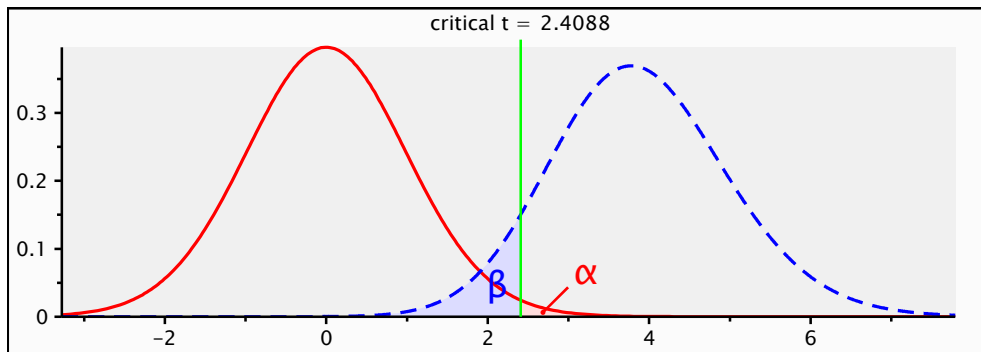


Abb. 7.18: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PR_1 (rot) vs. PT_1 (blau) bei einseitiger Testung (Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben) und $\alpha=1\%$, $df=46,7464829$

Nachfolgend werden die oben aufgestellten Hypothesen für das Verhältnis von Objektiv-Kurzska und Standard-Skala überprüft.

Tab. 7.54: Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben ($PR_{1(KS)}$ vs. PT_1)

	N	Mittlerer Rang	Rangsumme
Negative Ränge ^a	3	10,50	31,50 ^d
Positive Ränge ^b	17	10,50	178,50 ^e
Bindungen ^c	30	---	---
Gesamt	50	---	---
Z^f		-3,130	
Exakte Signifikanz		,001**	
φ^g		,443	
$1-\beta$,915 ^h		,741 ⁱ

^a Standard < Objektiv

^b Standard > Objektiv

^c Standard = Objektiv

^d W⁻

^e W⁺

^f z-verteilte Prüfgröße, basierend auf negativen Rängen

^g Effektstärke bzw. Effektgröße: $\varphi = \sqrt{\frac{z^2}{N}} = \frac{z}{\sqrt{N}}$ mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt)

(Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

^h post-hoc-Teststärke („observed power“) für $\alpha=5\%$, $df=46,7464829$: Wahrscheinlichkeit, mit der zu Gunsten der H_1 entschieden wird, wenn ein Unterschied besteht bzw. die H_1 gilt (Bortz (2005), S. 123; G* Power 3.1 manual (2007), S. 52ff)

ⁱ post-hoc-Teststärke („observed power“) für $\alpha=1\%$, $df=46,7464829$

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

(Quelle: eigene Berechnungen)

Auch der Median der Objektiv-Kurzska nimmt den Wert 3 an und unterscheidet sich somit augenscheinlich nicht vom Median der Standard-Skala. Der Wilcoxon-Test ergibt allerdings, dass in 3 Fällen der Vertrauenswert auf Basis der Objektiv-Kurzska höher als der Vertrauenswert auf Basis der Standard-Skala liegt, während in 23 Fällen das per Standard-Skala erfasste Vertrauen höher als das per Objektiv-Kurzska erfasste Vertrauen ist. In 30 Fällen ergibt sich keine Differenz zwischen den beiden Messvarianten. Das bedeutet: Es ergeben sich mehr Paardifferenzen, die in Richtung der Gültigkeit der H_1 weisen.

Die auf negativen Rängen basierende z-verteilte Prüfgröße zeigt ein hochsignifikantes Ergebnis an. Die Differenz zwischen den beiden Messvarianten ist damit als überzufällig zu bezeichnen und folgt der durch die Alternativhypothese vorgegebenen Richtung. Demnach sollte die H_0 abgelehnt und die H_1 beibehalten werden.

7. Hauptstudie

Zur Bestimmung der praktischen Bedeutsamkeit dieses Ergebnisses wurde auch für den hier vorgenommenen Vergleich Effektstärke und post-hoc-Teststärke berechnet. Die Effektstärke erreicht dabei einen Wert von $\varphi = .443$, was einem moderaten Effekt ($\omega \geq .30$) entspricht. Der Unterschied zwischen der Standard-Skala und der Objektiv-Kurzskala kann somit als praktisch relevant eingestuft werden. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 270, S. 633; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Nimmt man den empirischen z-Wert ($z = -3,130$) als Erwartungswert der Alternativhypothesenverteilung und ein 5%-Signifikanzniveau an, so ergibt sich eine post-hoc-Teststärke von $1 - \beta = .915$. Die Wahrscheinlichkeit, dass es sich bei der Entscheidung, die H_1 beizubehalten, um die richtige Entscheidung handelt, ist damit sehr hoch. (Bühner/Ziegler (2009), S. 191; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f) Testet man auf einem 1%-Signifikanzniveau, so sinkt die post-hoc-Teststärke ($1 - \beta = .741$) und sie befindet sich nicht mehr in einem tolerierbaren Bereich. Die H_1 kann bei $\alpha = 1\%$ nicht mehr als statistisch abgesichert angesehen werden ($1 - \beta < .80$). (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Cohen (1988), S. 4; Cohen (1992), S. 156)

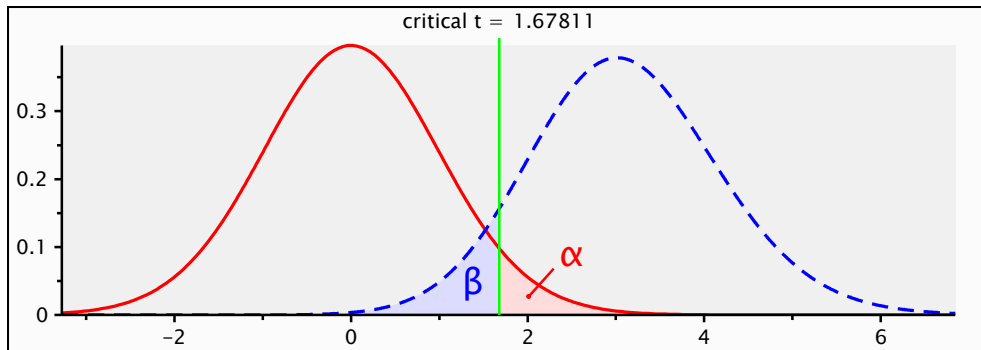


Abb. 7.19: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für $PR_{1(KS)}$ (rot) vs. PT_1 (blau) bei einseitiger Testung (Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben) und $\alpha = 5\%$, $df = 46,7464829$

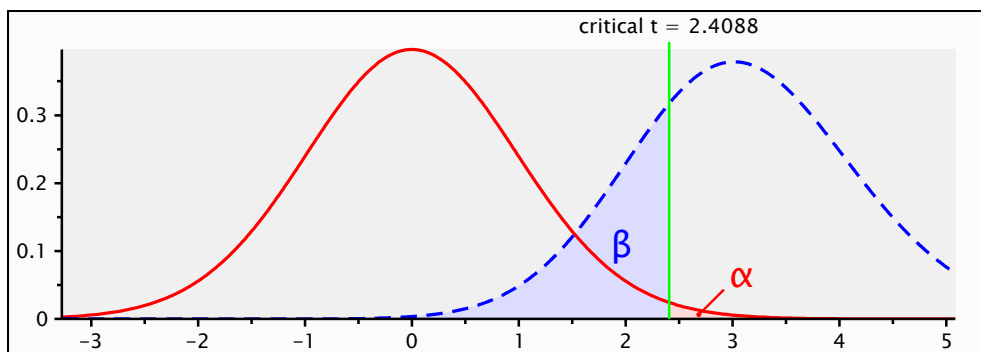


Abb. 7.20: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für $PR_{1(KS)}$ (rot) vs. PT_1 (blau) bei einseitiger Testung (Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben) und $\alpha = 1\%$, $df = 46,7464829$

7.2.6.7.1.2 Gruppe 2: Standard/Objektiv (PR_2 vs. PT_2)

Die zweite Versuchsgruppe kehrt im Vergleich zur ersten Versuchsgruppe die Bearbeitungsreihenfolge der Vertrauensskalen um. So bearbeiteten die Probanden innerhalb dieser Gruppe zunächst die Standard-Skala und im Anschluss die Objektiv-Skala.

7.2.6.7.1.2.1 Parametrische Signifikanzprüfung

Auch in der zweiten Versuchsgruppe soll nun anhand des T-Tests für abhängige Stichproben geprüft werden, ob sich das Antwortverhalten der Probanden je nach Messvariante signifikant unterscheidet bzw. ob die Probanden auf der Objektiv-Skala ein signifikant niedrigeres Vertrauen angeben als auf der Standard-Skala. Entsprechend der Bearbeitungsreihenfolge stellt die Standard-Variante nun Messzeitpunkt 1 dar (PR_2), während die Objektiv-Variante Messzeitpunkt 2 (PT_2) repräsentiert. Es ergeben sich hieraus die folgenden Hypothesen:

H_0 : Die Mittelwerte der Standard-Skala (PR_2) zur Erfassung Generalisierten Vertrauens und der Objektiv-Skala bzw. -Kurzskala (PT_2) zur Erfassung Generalisierten Vertrauens fallen innerhalb der Population gleich aus bzw. die Differenz der Messwertpaare in der Grundgesamtheit ist gleich null. Abweichungen zwischen den Mittelwerten von Objektiv-Skala bzw. -Kurzskala und Standard-Skala sind rein zufällig.

$$\mu_{PR_2} = \mu_{PT_2} \text{ bzw. } \mu_d = 0$$

H_1 : Der Mittelwert der Standard-Skala (PR_2) zur Erfassung Generalisierten Vertrauens fällt innerhalb der Population höher aus als der jeweilige Mittelwert der Objektiv-Skala bzw. -Kurzskala (PT_2) zur Erfassung Generalisierten Vertrauens bzw. die Differenz der Messwertpaare innerhalb der Grundgesamtheit ist größer null.

$$\mu_{PR_2} > \mu_{PT_2} \text{ oder } \mu_d > 0$$

Tab. 7.55: Parametrische Verfahren (T-Test für abhängige Stichproben)

Within-subject-Design	
PR_2 vs. PT_2	$T(49)=3,857$ (einseitig); $p<,01$; $g=,545$; $1-\beta=,922$
PR_2 vs. $PT_{2(KS)}$	$T(49)=2,214$ (einseitig); $p<,01$; $g=,313$; $1-\beta=,705$

$T(df)$: empirischem t-Wert inklusive Freiheitsgrade
(einseitig) bzw. (zweiseitig): Testrichtung
* $p<,05$ (einseitig), ** $p<,01$ (einseitig)
n.s.: nicht signifikant ($p>,05$ (zweiseitig))
g: Effektstärke mit $d = .20 \rightarrow$ *kleiner Effekt*, $d = .50 \rightarrow$ *mittlerer Effekt*, $d = .80 \rightarrow$ *großer Effekt* (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)
 $1-\beta$: Teststärke
Vgl. Bühner/Ziegler (2009), S. 247
(Quelle: eigene Berechnungen)

Sowohl der Unterschied zwischen den Messergebnissen der Standard- und der Objektiv-Skala als auch der Unterschied zwischen der Standard-Skala und der Objektiv-Kurzskala sind als überzufällig zu bewerten. Allerdings kann ausschließlich das signifikante Ergebnis des Kontrastes von PR_2 vs. PT_2 statistisch abgesichert werden. Die Wahrscheinlichkeit sich im Falle eines realen Unterschiedes zwischen den Messergebnissen für die Alternativhypothese zu entscheiden, ist hier somit als ausreichend hoch einzustufen. Im Gegensatz dazu liegt die Teststärke für den Kontrast zwischen PR_2 vs. $PT_{2(KS)}$ unterhalb des Schwellenwertes von $1-\beta=.80$. Die Wahrscheinlichkeit, rechtmäßig zugunsten der Alternativhypothese zu entscheiden, ist hier vermindert. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärke zeigt je nach Paarung einen schwachen bis mittleren Effekt an. Die Differenz zwischen den Messergebnissen scheint demnach in beiden Fällen praktisch relevant zu sein, fällt allerdings für den Kontrast von Standard- und Objektiv-Skala größer aus als für den Kontrast von

7. Hauptstudie

Standard-Skala und Objektiv-Kurzskala. In beiden Fällen ist im 95%-Konfidenzintervall der Differenz der Wert 0 nicht enthalten, was den Eindruck bestätigt, dass zwischen den Messmethoden ein Unterschied bestehen muss. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 187, S. 245, S. 250; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 373f)

Für beide Kontraste wird die H_0 abgelehnt und die H_1 beibehalten. Ausgehend von den Mittelwerten scheint die Differenz im Antwortverhalten zwischen den Skalen der Vorhersage durch die Alternativhypothese zu entsprechen. Die Probanden stellen sich auf der Standard-Skala signifikant vertrauensvoller dar als auf den Objektiv-Skalen. Für den Unterschied von Standard-Skala und Objektiv-Kurzskala ist allerdings zu bemerken, dass aufgrund der geringen Teststärke die Gefahr eines falsch-signifikanten Ergebnisses besteht.

7.2.6.7.1.2.2 Nonparametrische Signifikanzprüfung

Den Empfehlungen von BORTZ/LIENERT/BOEHNKE folgend sollen anschließend die Resultate der parametrischen Verfahren mit Hilfe verteilungsfreier Verfahren validiert werden. Für die zweite Versuchsgruppe soll daher anhand des Vorzeichen-Rang-Test von Wilcoxon geprüft werden, ob sich das Antwortverhalten der Probanden je nach Messvariante signifikant unterscheidet bzw. ob die Probanden auf der Objektiv-Skala ein signifikant niedrigeres Vertrauen angeben als auf der Standard-Skala. Entsprechend der Bearbeitungsreihenfolge stellt die Standard-Variante nun Messzeitpunkt 1 dar (PR_2), während die Objektiv-Variante Messzeitpunkt 2 (PT_2) repräsentiert. Es ergeben sich hieraus die folgenden Hypothesen:

H_0 : Die Mediane der Objektiv-Skala bzw. -Kurzskala (PT_2) zur Erfassung Generalisierten Vertrauens und der Standard-Skala (PR_2) zur Erfassung Generalisierten Vertrauens fallen in der Grundgesamtheit gleich aus bzw. die Rangsummen der beiden Messvarianten sind in der Population gleich. Unterschiede zwischen der Objektiv-Skala bzw. -Kurzskala und der Standard-Skala sind rein zufällig.

$$W^+ = W^- \text{ bzw. } \tilde{x}_{PR_2} = \tilde{x}_{PT_2}$$

H_1 : Der Median der Standard-Messung (PR_2) fällt innerhalb der Grundgesamtheit höher aus als der Median der Objektiv-Skala bzw. -Kurzskala (PT_2). Zwischen den beiden Messvarianten treten negative Ränge mit größerer Häufigkeit auf als positive Ränge.

$$W^+ < W^- \text{ bzw. } \tilde{x}_{PR} > \tilde{x}_{PT}$$

Tab. 7.56: Nonparametrische Verfahren (Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben)

Within-subject-Design	
PR₂ vs. PT₂	z=-3,409 (einseitig); p<.01; φ =.482; 1- β =.818
PR₂ vs. PT_{2(KS)}	z=-2,132 (einseitig); p<.01; φ =.302; 1- β =.660

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant (p>.05 (zweiseitig))

*p<.05 (einseitig), **p<.01 (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f;

Cohen (1988), S. 223ff)

1- β : Teststärke

vgl. Bühner/Ziegler (2009), S. 278

(Quelle: eigene Berechnungen)

Auch im Rahmen der nonparametrischen Verfahren ergibt sich für beide Paarungen ein überzufälliger Unterschied bei den Messergebnissen. Während das signifikante Ergebnis für den Kontrast von Standard- und Objektiv-Skala statistisch abgesichert werden kann, liegt die Teststärke bei der Gegenüberstellung von Standard-Skala und Objektiv-Kurzskala unterhalb der per Konvention festgelegten Schwelle von $1-\beta=.80$. Im ersten Fall ist die Wahrscheinlichkeit, die H_1 irrtümlich abzulehnen, damit als eher gering zu bewerten bzw. die Wahrscheinlichkeit, sich richtigerweise für die Alternativhypothese zu entscheiden, ist hoch. Im zweiten Fall ist die Wahrscheinlichkeit, eine korrekte Alternativhypothese anzunehmen, dagegen reduziert. (Bortz (2005), S. 122; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärke bewegt sich für beide Skalenpaarungen im Bereich für einen mittleren Effekt ($.30 \leq \omega \leq .50$). Die Differenz zwischen den Messergebnissen der verschiedenen Vertrauensskalen ist demnach als praktisch relevant zu bewerten. Im Falle des Kontrastes von PR_2 vs. $PT_{2(KS)}$ stellt dies außerdem einen Hinweis auf ein korrekt signifikantes Ergebnis dar. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 270, S. 633; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Für den Unterschied zwischen Standard- und Objektiv-Skala zeigt der Wilcoxon-Test, dass in 21 Fällen das durch die Objektiv-Skala abgebildete Vertrauen niedriger ist als das mit Hilfe der Standard-Skala ermittelte Vertrauen. Gleichzeitig ergibt sich lediglich für 4 Fälle die umgekehrte Wechselbeziehung. In 25 Fällen ergibt sich kein Unterschied zwischen den beiden Messvarianten. Die Rangsumme für negative Ränge ist dementsprechend höher als die für positive Ränge, was auf die Gültigkeit der H_1 hinweist.

Betrachtet man das Verhältnis der negativen und positiven Ränge für die Kontrastierung von Standard-Skala und Objektiv-Kurzskala wird sichtbar, dass die Rangsumme der negativen Ränge größer ist als die Rangsumme der positiven Ränge. In 16 Fällen liegt der per Objektiv-Kurzskala ermittelte Vertrauenswert unterhalb des per Standard-Skala ermittelten Vertrauenswertes. Lediglich in 6 Fällen kehrt sich diese Beziehung um. 28 Fälle weisen keinen Unterschied zwischen Standard-Skala und Objektiv-Kurzskala auf.

Ausgehend von der z-verteilten Prüfgröße ergibt sich aus der Differenz der Rangsummen ein signifikantes Ergebnis, wobei die Richtung des Unterschieds dem Postulat der Alternativhypothese entspricht. Demnach sollte in beiden Fällen die H_1 beibehalten und die H_0 abgelehnt werden. Auf den Objektiv-Skalen geben die Probanden somit einen niedrigeren Vertrauenslevel an als auf der Standard-Skala.

7.2.6.7.1.3 Zusammenfassung

Zunächst ist festzustellen, dass die durchschnittliche Differenz zwischen den Summenscores von Objektiv- und Standard-Skala bzw. Objektiv-Kurzskala und Standard-Skala weder in *Versuchsgruppe 1* noch in *Versuchsgruppe 2* einen halben Skalenpunkt überschreitet. Die Diskrepanz zwischen den Messergebnissen ist demnach verhältnismäßig klein, sodass es kaum einen Unterschied zu machen scheint, ob man zur Erfassung Generalisierten Vertrauens die im Zuge der Pilotstudie entwickelten Objektiv-Skalen oder die Standard-Skala bzw. *Trust in People-Scale* verwendet. Trotz allem handelt es sich bei der gefundenen Differenz um einen überzufälligen Unterschied, was eine praktische Bedeutsamkeit vermuten lässt bzw. anzeigt.

Ein weiterer interessanter Befund ergibt sich aus der Tatsache, dass der Unterschied zwischen Objektiv-Skala und Standard-Skala größer ausfällt als der Unterschied zwischen Objektiv-Kurzskala und Standard-Skala. Im Zusammenhang mit den Ergebnissen für die Konstruktvalidität lässt sich vermuten, dass die Objektiv-Kurzskala eine stärker ausgeprägte inhaltliche Kongruenz mit der Standard-Skala aufweist als die Objektiv-Skala. So ergaben sich im Rahmen der Bestimmung der Konstruktvalidität zwischen Objektiv-Kurzskala und Standard-Skala höhere Korrelationen als zwischen Objektiv-Skala und Standard-Skala. Dieses Ergebnis scheint mit der Differenz zwischen den Messergebnissen zu korrespondieren. Geht man davon aus, dass es sich bei den Ergebnissen der Standard-Skala um eine valide Abbildung Generalisierten Vertrauens handelt, so scheint die Objektiv-Kurzskala ein besseres Objektiv-Pendant zur Standard-Skala darzustellen als die längere Objektiv-Skala.

Sowohl die parametrischen als auch die nonparametrischen Verfahren ergeben in beiden Versuchsgruppen einen jeweils hochsignifikanten Unterschied zwischen den kontrastierten Messverfahren. Ausgehend von der Paralleltestempfehlung durch BORTZ/LIENERT/BOEHNKE kann somit davon ausgegangen werden, dass zwischen dem Antwortverhalten auf den Objektiv-Skalen und dem Antwortverhalten auf der Standard-Skala tatsächlich ein überzufälliger Unterschied besteht. Diese Interpretation wird zusätzlich dadurch gestützt, dass die Teststärken der Signifikanzprüfungen in den meisten Fällen entweder oberhalb des Schwellenwertes oder knapp unterhalb des Schwellenwertes liegen.

Dabei ist hervorzuheben, dass sich der beschriebene Unterschied unabhängig von der Präsentationsreihenfolge der Vertrauensskalen ergibt. So stellen sich die Probanden in beiden Versuchsgruppen auf der Standard-Skala vertrauensvoller dar als auf den Objektiv-Skalen. Ob diese Diskrepanz darauf zurückzuführen ist, dass die Probanden auf den Objektiv-Skalen ehrlicher antworten als auf der Standard-Skala oder die umgekehrte Relation gültig ist, kann an dieser Stelle allerdings nicht zwingend festgestellt werden.

Zusätzlich ist hervorzuheben, dass es sich bei den *Versuchsgruppen 1* und *2* jeweils um Versuchsanordnungen im *within-subject*-Design handelt. Die Probanden haben somit sowohl die Objektiv-Skalen als auch die Standard-Skalen bearbeitet. Demnach haben jeweils die gleichen Personen auf den verschiedenen Vertrauensskalen unterschiedlich geantwortet. Für den gefundenen Unterschied zwischen den Messergebnissen der Vertrauensskalen kann mit sehr hoher Wahrscheinlichkeit dementsprechend nur die Art und Weise der Abfrage verantwortlich sein. Einflüsse, die aus den Eigenschaften der Befragten resultieren könnten, wurden durch das Design konstant gehalten. In diesem Zusammenhang besteht allerdings die Möglichkeit, dass der Unterschied im Antwortverhalten der Probanden durch die geringe Augenscheinvalidität der Objektiv-Skalen bezüglich der Zieldimension „Generalisiertes Vertrauen“ und daraus möglicherweise resultierender unterschiedlicher Interpretationen der *demand characteristics* dieser Skalen entsteht.

7.2.6.7.2 Between-subject-Design

Im folgenden Abschnitt soll überprüft werden, ob sich zwischen den Versuchsgruppen, die in einem *between-subject*-Design angeordnet sind bzw. angeordnet werden können, überzufällige Unterschiede bezüglich der Messergebnisse zeigen. Genuin als *between-subject*-Design angelegt war im Rahmen der experimentellen Anordnung die Kontrastierung von *Gruppe 3: Objektiv (PT₃)* und *Gruppe 4: Standard (PT₄)*. Daneben können aber auch die nachfolgend aufgeführten Gegenüberstellungen als *between-subject*-Designs interpretiert werden:

Gruppe 1: Objektiv (PR₁) vs. Gruppe 2: Standard (PR₂)

Gruppe 1: Objektiv (PR₁) vs. Gruppe 4: Standard (PT₄)

Gruppe 2: Standard (PR₂) vs. Gruppe 3: Objektiv (PT₃)

Für die genannten Kontraste kann angenommen werden, dass die Probanden die jeweiligen Einzelmessungen als *One-Shot-Case-Study* bzw. *ex-post-facto-Design* wahrgenommen haben. Diese Annahme muss dabei insbesondere für die Pretests aus den *Versuchsgruppen 1* und *2* Gültigkeit besitzen. Die Posttests in den *Versuchsgruppen 1* und *2* können dagegen nicht als Einzelmessung interpretiert werden, da im Vorhinein bereits die Pretests präsentiert wurden.

Nachteil des *between-subject*-Design ist es, dass etwaige Unterschiede zwischen den Versuchsgruppen nicht zweifelsfrei dem Treatment zugeschrieben werden können, sondern auch Unterschiede, die auf die personelle Zusammensetzung der unterschiedlichen Gruppen zurückgehen, für die Differenz verantwortlich sein können. So werden im Rahmen des *between-subject*-Designs anders als beim *within-subject*-Design voneinander verschiedene Stichproben miteinander verglichen.

7.2.6.7.2.1 Gruppe 3: Objektiv (PT₃) vs. Gruppe 4: Standard (PT₄)

Nachfolgend soll mit Hilfe von parametrischen und nonparametrischen Signifikanztests für unabhängige Stichproben geprüft werden, ob sich die Messergebnisse der Vertrauensskalen aus *Gruppe 3: Objektiv (PT₃)* von denen aus *Gruppe 4: Standard (PT₄)* signifikant unterscheiden. Die Besonderheit von *Gruppe 3: Objektiv (PT₃)* und *Gruppe 4: Standard (PT₄)* besteht dabei darin, dass innerhalb der Gruppen jeweils nur eine der Messvarianten zum Einsatz kam und somit Verzerrungen im Antwortverhalten, die durch etwaige *Pretest*-, *Reihenfolge*- oder *Carry-Over*-Effekte induziert sein könnten, ausgeschlossen werden können.

7.2.6.7.2.1.1 Parametrische Signifikanzprüfung

Zunächst kommt der T-Test für unabhängige Stichproben zur Prüfung des Kontrastes zwischen den beiden genannten Versuchsgruppen zum Einsatz. Es ergeben sich dann die folgenden Hypothesen:

H₀: Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PT₃) zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 3* unterscheiden sich jeweils nicht vom Mittelwert der Standard-Skala zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 4: Standard (PT₄)*. Die Differenz der Populationsmittelwerte ergibt null.

$$\mu_{PT3} = \mu_{PT4} \text{ oder } \mu_{PT3} - \mu_{PT4} = 0$$

H₁: Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PT₃) zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 3* fallen jeweils niedriger aus als der Mittelwert der Standard-Skala zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 4: Standard (PT₄)*. Die Differenz der Populationsmittelwerte ist kleiner als null.

$$\mu_{PT3} < \mu_{PT4} \text{ oder } \mu_{PT3} - \mu_{PT4} < 0$$

Tab. 7.57: Levene-Test und T-Test für unabhängige Stichproben (PT₃ vs. PT₄)

	N	\bar{x}	s
Objektiv (PT ₃)	52	3,0192	,61006
Standard (PT ₄)	52	3,1923	,81742
F ^a		12,515	
Signifikanz		,001 ^{**}	
$\bar{x}_1 - \bar{x}_2$ ^b		-,17308	
$\hat{\sigma}_p$ ^c		,71374	
Konf _{95%} (Differenz) ^d		- ,45391	,10775
T ^e		-1,224	
Df ^f		94,361	
Signifikanz		,224 ^{n.s.}	
Hedges g/Cohens d _s ^g		,243	
1- β		,233 ^h	
β		,767 ⁱ	

^a F-verteilte Prüfgröße^b Mittlere Differenz^c gepoolte Standardabweichung: $\hat{\sigma}_p = \frac{s_1 + s_2}{2}$ (Bühner/Ziegler (2009), S. 266; Eid/Gollwitzer/Schmitt (2015), S. 337f)^d 95% Konfidenzintervall der Differenz; Links: Untere Grenze, Rechts: Obere Grenze^e t-verteilte Prüfgröße^f Freiheitsgrade (*degrees of freedom*) mit $df=n_1+n_2-2$ bei homogenen Varianzen bzw. $df_{\text{kor}} = \frac{(n_1-1) \cdot (n_2-1)}{(n_2-1) \cdot c^2 + (n_1-1) \cdot (1-c)^2}$ mit $c = \frac{\frac{\hat{\sigma}_1^2}{n_1}}{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}$ bei

heterogenen Varianzen (Eid/Gollwitzer/Schmitt (2015), S. 254, S. 333ff; Bortz (2005), S. 141; Bühner/Ziegler (2009), S. 261f; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 162ff)

^g Effektstärke bzw. Effektgröße: Hedges g bzw. Cohens d_s, mit g bzw. $d_s = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}_p}$; $d = .20 \rightarrow \text{kleiner Effekt}$, $d = .50 \rightarrow \text{mittlerer Effekt}$, $d = .80 \rightarrow \text{großer Effekt}$ (Eid/Gollwitzer/Schmitt (2015), S. 337f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 179, S. 258)^h post-hoc-Teststärke (*power*) für $\alpha=5\%$, $df=102$: Wahrscheinlichkeit, mit der zu Gunsten der H₁ entschieden wird, wenn ein Unterschied besteht bzw. die H₁ gilt (Bortz (2005), S. 123; Bühner/Ziegler (2009), S. 266f)ⁱ β -Fehler-Wahrscheinlichkeit für $\alpha=5\%$, $df=94,361$: „Die Irrtumswahrscheinlichkeit β ist die Wahrscheinlichkeit, mit der ein statistischer Test ein nicht-signifikantes Ergebnis ergibt, obwohl in Wirklichkeit die H₁ gilt.“ (Eid/Gollwitzer/Schmitt (2015), S. 224)* $p < .05$ (einseitig), ** $p < .01$ (einseitig)^{n.s.} nicht signifikant ($p > .05$ (zweiseitig))

(Quelle: eigene Berechnungen)

Vergleicht man die Mittelwerte der beiden Messvarianten miteinander, so ergibt sich, dass die Versuchspersonen auf der Objektiv-Variante im Durchschnitt ein geringeres Generalisiertes Vertrauen angeben als die Versuchspersonen, die die Standard-Variante bearbeitet hatten. Ein potentieller Unterschied im Antwortverhalten folgt somit der von der Alternativhypothese postulierten Richtung.

Ein weiterer Hinweis auf Unterschiede im Antwortverhalten ergibt sich aus dem signifikanten Ergebnis des Levene-Tests. So scheinen die Versuchspersonen in PT₄ differenzierter zu antworten als in PT₃.

Der t-verteilten Prüfgröße zur Folge ergibt sich kein signifikanter Unterschied im Antwortverhalten zwischen den Messverfahren zur Erfassung Generalisierten Vertrauens. Es wird die H₀ beibehalten und die H₁ abgelehnt.

Folgt man der Effektstärke, so ergibt sich mit $g/d_s = .243$ ein kleiner Effekt. Die Differenz zwischen PT₃ und PT₄ ist damit ungleich null und es scheint trotz des nicht signifikanten Ergebnisses ein praktisch bedeutsamer Kontrast im Antwortverhalten zu bestehen. Gegen diese Interpretation spricht allerdings, dass das 95%-Konfidenzintervall der Differenz den Wert null enthält. Bei der vorliegenden Konstellation der Werte liegt demnach auch eine Null Differenz zwischen den Mittelwerten im Bereich des Möglichen. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

Unter der Bedingung einer spezifischen Alternativhypothese mit $t = -1,224$ als Erwartungswerts und $\alpha = 5\%$ ergibt sich für die post-hoc-Teststärke ein Wert von $1 - \beta = .233$ und für die β -Fehler-Wahrscheinlichkeit $\beta = .767$. Ausgehend von der Teststärke hat die H₁ damit eine zu geringe

7. Hauptstudie

Chance erhalten sich zu bewahrheiten ($1-\beta < .80$), während gleichzeitig die Gefahr für ein falsch nicht-signifikantes Ergebnis oberhalb des zulässigen Grenzwertes von $\beta = .20$ liegt. Insgesamt ist damit das Resultat des T-Tests nicht eindeutig interpretierbar. Es scheinen zu wenige Daten vorzuliegen, als dass eine eindeutige Entscheidung zu Gunsten der H_1 gefällt werden könnte. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

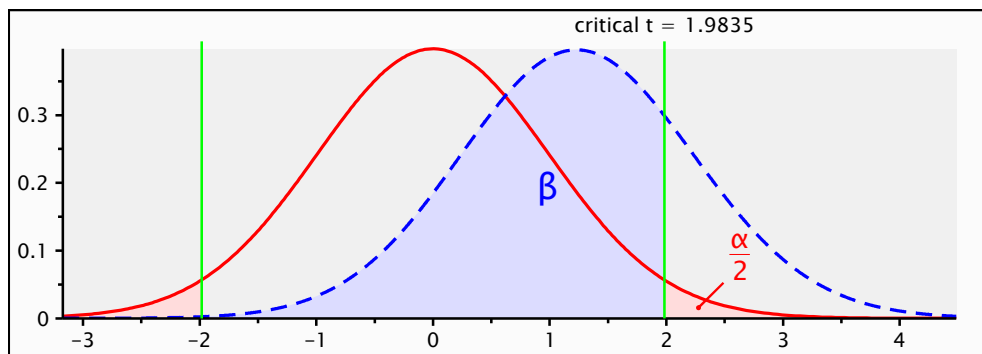


Abb. 7.21: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PT_3 (rot) vs. PT_4 (blau) bei zweiseitiger Testung (T-Test für unabhängige Stichproben) und $\alpha=5\%$, $df=102$

Nachfolgend soll auch der potentielle Unterschied zwischen der aus PT_3 hervorgehenden Objektiv-Kurzskala und der Standard-Skala aus PT_4 betrachtet werden.

Tab. 7.58: Levene-Test und T-Test für unabhängige Stichproben ($PT_{3(KS)}$ vs. PT_4)

	N	\bar{x}	s
Objektiv ($PT_{3(KS)}$)	52	3,0769	,58899
Standard (PT_4)	52	3,1923	,81742
F^a		12,573	
Signifikanz		,001**	
$\bar{x}_1 - \bar{x}_2^b$		-,11538	
$\hat{\sigma}_p^c$,70321	
Konf _{95%} (Differenz) ^d	-,39285		,16208
T^e		-,826	
Df ^f		92,713	
Signifikanz		,411 ^{n.s.}	
Hedges g/Cohens d_s^g		,164	
1- β		,132 ^h	
β		,868 ⁱ	

^a F-verteilte Prüfgröße

^b Mittlere Differenz

^c gepoolte Standardabweichung: $\hat{\sigma}_p = \frac{s_1 + s_2}{2}$ (Bühner/Ziegler (2009), S. 266; Eid/Gollwitzer/Schmitt (2015), S. 337f)

^d 95% Konfidenzintervall der Differenz; Links: Untere Grenze, Rechts: Obere Grenze

^e t-verteilte Prüfgröße

^f Freiheitsgrade (degrees of freedom) mit $df=n_1+n_2-2$ bei homogenen Varianzen bzw. $df_{korr} = \frac{(n_1-1) \cdot (n_2-1)}{(n_2-1) \cdot c^2 + (n_1-1) \cdot (1-c)^2}$ mit $c = \frac{\frac{\hat{\sigma}_1^2}{n_1}}{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}$ bei

heterogenen Varianzen (Eid/Gollwitzer/Schmitt (2015), S. 254, S. 333ff; Bortz (2005), S. 141; Bühner/Ziegler (2009), S. 261f; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 162ff)

^g Effektstärke bzw. Effektgröße: Hedges g bzw. Cohens d_s , mit g bzw. $d_s = \frac{|x_1 - x_2|}{\hat{\sigma}_p}$; $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 337f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 179, S. 258)

^h post-hoc-Teststärke (power) für $\alpha=5\%$, $df=102$: Wahrscheinlichkeit, mit der zu Gunsten der H_1 entschieden wird, wenn ein Unterschied besteht bzw. die H_1 gilt (Bortz (2005), S. 123; Bühner/Ziegler (2009), S. 266f)

ⁱ β -Fehler-Wahrscheinlichkeit für $\alpha=5\%$, $df=92,713$: „Die Irrtumswahrscheinlichkeit β ist die Wahrscheinlichkeit, mit der ein statistischer Test ein nicht-signifikantes Ergebnis ergibt, obwohl in Wirklichkeit die H_1 gilt.“ (Eid/Gollwitzer/Schmitt (2015), S. 224)

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

^{n.s.} nicht signifikant ($p > .05$ (zweiseitig))

(Quelle: eigene Berechnungen)

Der Augenscheinvergleich der Mittelwerte der Objektiv-Kurzskala und der Standard-Skala ergibt das gleiche Resultat wie in der vorherigen Analyse. Auf der Objektiv-Kurzskala erreichen die Versuchspersonen ein niedrigeres Generalisiertes Vertrauen als die Versuchspersonen auf der Standard-Skala. Ein potentieller Unterschied im Antwortverhalten würde somit entlang der Alternativhypothese verlaufen.

Auch aus dem Levene-Test ergibt sich ein Hinweis auf einen Unterschied im Antwortverhalten je nach bearbeiteter Skala. So scheinen die Probanden in PT₄ differenzierter zu antworten als die Probanden in PT₃.

Die t-verteilte Prüfgröße zeigt ein nicht signifikantes Ergebnis an. Es wird entsprechend die H_0 beibehalten und die H_1 abgelehnt.

Anhand der Effektstärke lässt sich ablesen, dass trotz nicht signifikantem Ergebnis ein minimaler Unterschied zwischen den Messvarianten bestehen muss. Mit $g/d_s=.164$ ist der Kontrast zwar ungleich null, liegt jedoch unterhalb des Schwellenwertes für einen kleinen Effekt. Allerdings überstreicht das 95%-Konfidenzintervall der Differenz den Wert null, sodass mit einer hohen Wahrscheinlichkeit angenommen werden kann, dass kein Unterschied zwischen den Mittelwerten der verschiedenen Messmethoden besteht. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

Unter der Bedingung des empirischen t-Wertes als Erwartungswert der Alternativhypotheseverteilung ergibt sich für die post-hoc-Teststärke ein Wert unterhalb des notwendigen Niveaus von $1-\beta=.80$. Die Wahrscheinlichkeit, zu Gunsten der H_1 zu entscheiden, wenn diese tatsächlich Gültigkeit besitzt, ist mit $1-\beta=.132$ stark reduziert. Gleichzeitig rangiert die β -Fehler-Wahrscheinlichkeit mit $\beta=.868$ weit oberhalb des Grenzwertes von $\beta=.20$. Ein falsch signifikantes Ergebnis kann demnach nicht ausgeschlossen werden. Allerdings ist eine eindeutige Interpretation aufgrund der geringen Teststärke nicht möglich. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

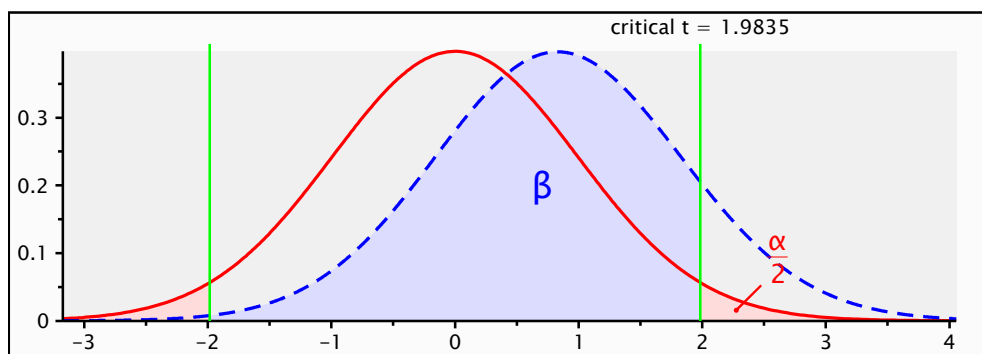


Abb. 7.22: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PT_{3(KS)} (rot) vs. PT₄ (blau) bei zweiseitiger Testung (T-Test für unabhängige Stichproben) und $\alpha=5\%$, $df=102$

7.2.6.7.2.1.2 Nonparametrische Signifikanzprüfung

Nachfolgend soll auch mit Hilfe des U-Tests geprüft werden, ob sich ein Unterschied zwischen *Gruppe 3: Objektiv* (PT₃) und *Gruppe 4: Standard* (PT₄) ergibt. Dabei werden die folgenden Hypothesen geprüft:

H₀: Die Mediane der Objektiv-Skala bzw. -Kurzskala (PT₃) zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 3: Objektiv* unterscheiden sich nicht vom Median der Standard-Skala zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 4: Standard* (PT₄). In der Grundgesamtheit liegen gleich viele Rangplatzüberschreitungen wie -unterschreitungen vor.

$$\tilde{x}_{PT3} = \tilde{x}_{PT4}$$

H₁: Die Mediane der Objektiv-Skala bzw. -Kurzskala (PT₃) zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 3: Objektiv* sind kleiner als der Median der Standard-Skala zur Erfassung Generalisierten Vertrauens des Posttests aus *Gruppe 4: Standard* (PT₄). In der Grundgesamtheit liegen mehr Rangplatzunterschreitungen als Rangplatzüberschreitungen vor.

$$\tilde{x}_{PT} < \tilde{x}_{PT}$$

Tab. 7.59: Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben (PT₃ vs. PT₄)

	N	Mittlerer Rang ^a	Rangsumme
Objektiv (PT ₃)	52	49,07	2551,50
Standard (PT ₄)	52	55,93	2908,50
Gesamt	104	---	---
Mann-Whitney-U ^b	1173,50		
Wilcoxon-W ^c	2551,50		
Z ^d	-1,264		
Exakte Signifikanz	,207 ^{n.s.}		
φ ^e	,124		
1-β ^f	,094		
β	,906 ^g		

^a durchschnittlicher Rangplatz

^b Prüfgröße U; für größere Stichproben (n>20) nähert sich U einer Normalverteilung an und kann näherungsweise durch einen z-Wert ersetzt werden (Eid/Gollwitzer/Schmitt (2015), S. 345f; Bühner/Ziegler (2009), S. 284)

^c Prüfgröße W

^d z-verteilte Prüfgröße, basierend auf Prüfgröße U (Eid/Gollwitzer/Schmitt (2015), S. 345f; Bühner/Ziegler (2009), S. 284)

^e Effektstärke bzw. Effektgröße: $\varphi = \sqrt{\frac{z^2}{N}} = \frac{z}{\sqrt{N}}$ mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

^f post-hoc-Teststärke („observed power“) für $\alpha=5\%$, $df=97,3126845$: Wahrscheinlichkeit, mit der zu Gunsten der H₁ entschieden wird, wenn ein Unterschied besteht bzw. die H₁ gilt (Bortz (2005), S. 123; G* Power 3.1 manual (2017), S. 56ff)

^g β -Fehler-Wahrscheinlichkeit für $\alpha=5\%$, $df=97,3126845$: „Die Irrtumswahrscheinlichkeit β ist die Wahrscheinlichkeit, mit der ein statistischer Test ein nicht-signifikantes Ergebnis ergibt, obwohl in Wirklichkeit die H₁ gilt.“ (Eid/Gollwitzer/Schmitt (2015), S. 224)

^{n.s.} nicht signifikant (p>.05 (zweiseitig))

(Quelle: eigene Berechnungen)

Wie schon bereits bei den Medianen der Messvarianten innerhalb von *Gruppe 1: Objektiv/Standard* und *Gruppe 2: Standard/Objektiv* beobachtet, unterscheiden sich auch die Mediane von *Gruppe 3: Objektiv* und *Gruppe 4: Standard* nicht voneinander, sondern nehmen jeweils den Wert 3 an. Ausgehend von der in Augenscheinahme der Mediane kann somit auf keinen Unterschied zwischen den beiden Messvarianten geschlossen werden.

Betrachtet man die mittleren Ränge der beiden unabhängigen Stichproben, so lässt sich feststellen, dass die Probanden in *Gruppe 3: Objektiv* im Durchschnitt niedrigere Ränge als die

7. Hauptstudie

Probanden der *Gruppe 4: Standard* haben. Dies lässt darauf schließen, dass die in *Gruppe 3: Objektiv* erreichten Vertrauenswerte im Mittel geringer sind als die in *Gruppe 4: Standard* erreichten Vertrauenswerte. (Bühner/Ziegler (2009), S. 287) Die Richtung des Unterschieds entspricht damit der Alternativhypothese.

Die z-verteilte Prüfgröße zeigt allerdings, dass der Unterschied zwischen den Rangsummen nicht ausreichend groß und dementsprechend nicht signifikant wird. Ausgehend von diesem Ergebnis wird somit die H_1 abgelehnt und die H_0 beibehalten.

Die Effektstärke zeigt mit $\varphi = .124$ einen schwachen Effekt an ($\omega \geq .10$). Die Parameterdifferenz der beiden unabhängigen Stichproben *Gruppe 3: Objektiv* und *Gruppe 4: Standard* kann somit als gering bezeichnet werden. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 269f, S. 633; Cohen (1988), S. 9f, S. 223ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Nimmt man den empirischen z-Wert ($z = -1,264$) als Erwartungswert für die Verteilung der Werte unter der Bedingung der Gültigkeit der Alternativhypothese an, so ergibt sich eine post-hoc-Teststärke von $1 - \beta = .094$. Die Wahrscheinlichkeit, die Nullhypothese abzulehnen, liegt damit weit unterhalb des notwendigen Niveaus von 80% bzw. die Chance auf ein signifikantes Ergebnis ist als sehr gering zu bewerten. Die β -Fehler-Wahrscheinlichkeit beträgt 90,6% und liegt damit weit oberhalb der Toleranz von $\beta = .20$. Es ist daher anzunehmen, dass es sich bei dem hier gefundenen Ergebnis um ein falsch nicht signifikantes Ergebnis handelt bzw. ein Fehler 2. Art eingetreten ist. Mit großer Wahrscheinlichkeit wurde die H_1 zu Unrecht abgelehnt bzw. die Datenbasis des Vergleichs ist zu gering.⁸⁵ Dies ist insbesondere vor dem Hintergrund, dass die Effektstärke von null abweicht, zu vermuten. Würde die H_0 Gültigkeit besitzen bzw. überhaupt kein Unterschied zwischen den beiden Messmethoden bestehen, so müsste $\varphi = 0$ gelten. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff, S. 10; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

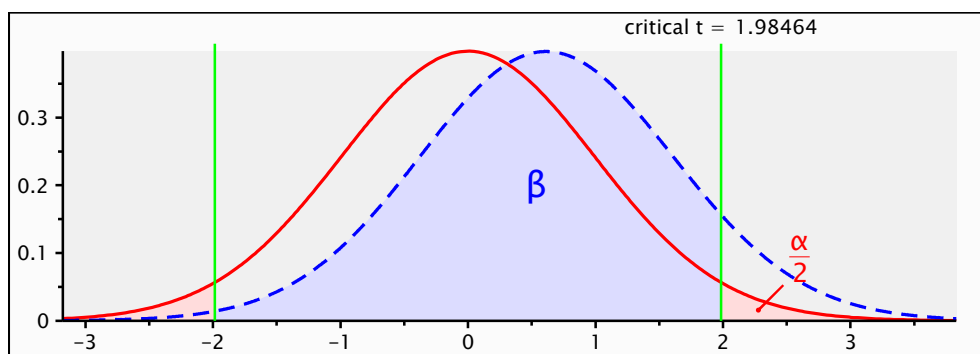


Abb. 7.23: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für PT_3 (rot) vs. PT_4 (blau) bei zweiseitiger Testung (Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben) und $\alpha = 5\%$, $df = 97, 3126845$

Es soll nun auch geprüft werden, ob sich die Standard-Skala und die Objektiv-Kurzskala bei der Kontrastierung von PT_3 und PT_4 voneinander unterscheiden.

⁸⁵ „Bei einem statistisch *nicht-signifikanten Ergebnis* muss immer auch die *Teststärke* bestimmt werden. Nur wenn die Teststärke mindestens 80% beträgt ($1 - \beta \geq .80$), kann das nicht-signifikante Ergebnis zum Anlass genommen werden, die H_0 anzunehmen. Liegt die Teststärke unter 80% bzw. umgekehrt die Beta-Fehler-Wahrscheinlichkeit über 20% ($\beta > .20$), so ist das Ergebnis nicht eindeutig interpretierbar. Der Signifikanztest basiert dann offensichtlich auf einer zu geringen Datenbasis, um eine Entscheidung H_1 über das Hypothesenpaar zu treffen.“ (Bortz/Döring (2016), S. 670)

Tab. 7.60: Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben (PT_{3(KS)} vs. PT₄)

	N	Mittlerer Rang ^a	Rangsumme
Objektiv (PT ₃)	52	50,04	2602,00
Standard (PT ₄)	52	54,96	2858,00
Gesamt	104	---	---
Mann-Whitney-U ^b	1224,00		
Wilcoxon-W ^c	2602,00		
Z ^d	-,911		
Exakte Signifikanz	,373 ^{n.s.}		
φ ^e	,089		
1- β ^f	,072		
β	,928 ^g		

^a durchschnittlicher Rangplatz^b Prüfgröße U; für größere Stichproben (n>20) nähert sich U einer Normalverteilung an und kann näherungsweise durch einen z-Wert ersetzt werden (Eid/Gollwitzer/Schmitt (2015), S. 345f; Bühner/Ziegler (2009), S. 284)^c Prüfgröße W^d z-verteilte Prüfgröße, basierend auf Prüfgröße U (Eid/Gollwitzer/Schmitt (2015), S. 345f; Bühner/Ziegler (2009), S. 284)^e Effektstärke bzw. Effektgröße: $\varphi = \sqrt{\frac{z^2}{N}} = \frac{z}{\sqrt{N}}$ mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)^f post-hoc-Teststärke („observed power“) für $\alpha=5\%$, $df=97,3126845$: Wahrscheinlichkeit, mit der zu Gunsten der H_1 entschieden wird, wenn ein Unterschied besteht bzw. die H_1 gilt (Bortz (2005), S. 123; G* Power 3.1 manual (2017), S. 56ff)^g β -Fehler-Wahrscheinlichkeit für $\alpha=5\%$, $df=97,3126845$: „Die Irrtumswahrscheinlichkeit β ist die Wahrscheinlichkeit, mit der ein statistischer Test ein nicht-signifikantes Ergebnis ergibt, obwohl in Wirklichkeit die H_1 gilt.“ (Eid/Gollwitzer/Schmitt (2015), S. 224)^{n.s.} nicht signifikant ($p>.05$ (zweiseitig))

(Quelle: eigene Berechnungen)

Auch der Median der Objektiv-Kurzskala aus *Gruppe 3: Objektiv* unterscheidet sich per Augenschein nicht vom Median der Standard-Skala aus *Gruppe 4: Standard*. Beide nehmen den Wert 3 an. Ausgehend von der in Augenscheinnahme der Mediane kann somit auf keinen Unterschied zwischen den beiden Messvarianten zur Erfassung Generalisierten Vertrauens geschlossen werden.

Ein Vergleich der mittleren Ränge ergibt, dass die Probanden innerhalb der *Gruppe 3: Objektiv* im Mittel niedrigere Ränge erreichen als die Probanden innerhalb von *Gruppe 4: Standard*. Es kann daraus geschlossen werden, dass in *Gruppe 3: Objektiv* durchschnittlich niedrigere Vertrauenswerte erreicht werden als in *Gruppe 4: Standard* bzw. ein potentieller Unterschied zwischen beiden Messmethoden der von der Alternativhypothese postulierten Richtung entspricht. (Bühner/Ziegler (2009), S. 287)

Aus der z-verteilten Prüfgröße lässt sich indes ablesen, dass der Unterschied zwischen den Rangsummen nicht ausreichend groß ist und sich kein signifikantes Ergebnis einstellt. Es sollte demnach die H_0 beibehalten und die H_1 abgelehnt werden.

Die Stärke des Effektes beträgt $\varphi = .089$ und ist als sehr gering zu bezeichnen. Da die Effektgröße allerdings von null abweicht, kann angenommen werden, dass innerhalb der Population trotz allem eine geringe Differenz zwischen den verschiedenen Varianten zur Erfassung Generalisierten Vertrauens besteht. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 269f, S. 633; Cohen (1988), S. 9f, S. 223ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Unter der Bedingung einer spezifischen Alternativhypothese mit dem empirischen z-Wert ($z=-.911$) als Erwartungswert und einer Signifikanzschwelle von $\alpha=5\%$ ergibt sich für die post-hoc-Teststärke $1-\beta=.072$. Die Chance für die Beibehaltung einer korrekten H_1 liegt damit weit unterhalb des angestrebten Levels von 80%. Für die β -Fehler-Wahrscheinlichkeit ergibt sich entsprechend $\beta=.928$. Die Gefahr eines β -Fehlers liegt somit weit oberhalb des akzeptierten

7. Hauptstudie

Niveaus von 20%. Die Teststärke ist zu gering, als dass das nicht-signifikante Ergebnis des Tests ohne Weiteres angenommen werden kann. Gleichzeitig ist das Ergebnis aber auch nicht eindeutig in Richtung der H_1 interpretierbar, da die zugrundeliegende Datenbasis des Tests zu gering erscheint. Alles in allem scheint allerdings die Möglichkeit eines korrekt nicht signifikanten Ergebnisses sehr wahrscheinlich. Dies ist vor allem deshalb zu vermuten, weil die Effektstärke einen Wert unterhalb der Schwelle für einen schwachen Effekt annimmt ($\omega \leq .10$). Als Argument für ein falsch nicht signifikantes Ergebnis lässt sich indes anführen, dass die Effektstärke einen Wert ungleich null annimmt. Dies bedeutet, dass nicht keine Differenz im Antwortverhalten auf den verschiedenen Vertrauensskalen vorliegt. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 670; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff, S. 10; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

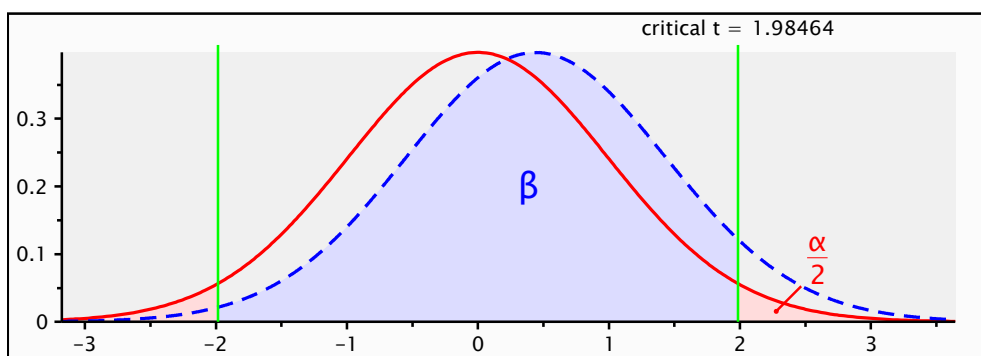


Abb. 7.24: Ablehnungsbereiche von H_1 (β -Fehler) bzw. H_0 (α -Fehler) für $PT_{3(KS)}$ (rot) vs. PT_4 (blau) bei einseitiger Testung (Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben) und $\alpha=5\%$, $df=97,3126845$

7.2.6.7.2.2 Gruppenübergreifende Kontrastierung (PR_1 vs. PR_2 , PR_1 vs. PT_4 und PR_2 vs. PT_3)

Wie bereits im Rahmen des Kapitels „Auswertungskonzept“ beschrieben, lässt sich auch anhand der Gegenüberstellung der Pretests aus den *Gruppen 1* und *2* die Differenz zwischen den beiden Messmethoden zur Erfassung Generalisierten Vertrauens extrahieren. Gleiches gilt für die Kontrastierung des Pretests aus *Versuchsgruppe 1* und des Posttests aus *Versuchsgruppe 4* sowie die Kontrastierung des Pretests aus *Gruppe 2* und des Posttests aus *Gruppe 3*. Die genannten Messungen enthalten jeweils keine Pretest-, Positions- oder Carry-over-Effekte, da ihnen keine Messung vorangegangen war. Sie stellen in diesem Sinne jeweils sogenannte *one-shot-case-studies* dar, die zur Bestimmung des Kontrastes der Messmethoden in einem *between-subject-Design* angeordnet werden können. Die Beziehung von PR_1 vs. PR_2 , PR_1 vs. PT_4 und PR_2 vs. PT_3 kann demnach analog zur Beziehung von PT_3 und PT_4 interpretiert werden.

7.2.6.7.2.2.1 Parametrische Signifikanzprüfung

Zunächst sollen die Unterschiede zwischen den Vertrauensmessungen mit Hilfe parametrischer Verfahren untersucht werden. Da es sich bei den zueinander in Beziehung gesetzten Versuchsgruppen um wechselseitig unabhängige Stichproben handelt, kommt zur Ermittlung des Unterschieds zwischen der Objektiv- und der Standard-Skala der T-Test für unabhängige

7. Hauptstudie

Stichproben zum Einsatz. Im Allgemeinen ergeben sich dabei die folgenden zu testenden Hypothesen:

H_0 : Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR_1 , PT_3) zur Erfassung Generalisierten Vertrauens unterscheiden sich jeweils nicht vom Mittelwert der Standard-Skala (PR_2 , PT_4) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwerte ergibt null.

$$\mu_{PR1} = \mu_{PR2} \text{ oder } \mu_{PR1} - \mu_{PR2} = 0$$

$$\mu_{PR1} = \mu_{PT4} \text{ oder } \mu_{PR1} - \mu_{PT4} = 0$$

$$\mu_{PR2} = \mu_{PT3} \text{ oder } \mu_{PR2} - \mu_{PT3} = 0$$

H_1 : Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR_1 , PT_3) zur Erfassung Generalisierten Vertrauens fallen jeweils niedriger aus als der Mittelwert der Standard-Skala (PR_2 , PT_4) zur Erfassung Generalisierten Vertrauens, bzw. der Mittelwert der Standard-Skala (PR_2 , PT_4) zur Erfassung Generalisierten Vertrauens fällt jeweils höher aus als die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR_1 , PT_3) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwerte ist kleiner bzw. größer als null.

$$\mu_{PR1} < \mu_{PR2} \text{ oder } \mu_{PR1} - \mu_{PR2} < 0$$

$$\mu_{PR1} < \mu_{PT4} \text{ oder } \mu_{PR1} - \mu_{PT4} < 0$$

$$\mu_{PR2} > \mu_{PT3} \text{ oder } \mu_{PR2} - \mu_{PT3} > 0$$

Zunächst soll der Vergleich von Objektiv- und Standard-Skala diskutiert werden. Hier ergab sich für jede der Gegenüberstellungen ein überzufälliger Unterschied zwischen den Messergebnissen, wobei allerdings nur das hochsignifikante Ergebnis der Kontrastierung von PR_1 und PR_2 statistisch abgesichert werden konnte. Für die übrigen Kontraste liegt die post-hoc-Teststärke unterhalb des notwendigen Niveaus von $1-\beta=.80$, sodass für diese Vergleiche die Wahrscheinlichkeit, rechtmäßig zugunsten der Alternativhypothese zu entscheiden, als reduziert angesehen werden muss. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Werte der Effektstärke zeigen je nach Paarungen einen kleinen bis moderaten Effekt an. Trotz der vom Betrag her geringen Differenz zwischen den Messergebnissen, scheint diese praktisch bedeutsam zu sein. Diese Beobachtung wird zusätzlich durch das jeweilige 95%-Konfidenzintervall der Differenz gestützt, dessen Wertebereich den Wert 0 nicht überspannt. Hieraus lässt sich schließen, dass mit hoher Wahrscheinlichkeit tatsächlich ein von null verschiedener Unterschied zwischen den Messverfahren existiert. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

In allen drei Fällen wird die H_0 abgelehnt und die H_1 beibehalten. Somit verläuft die Differenz im Antwortverhalten zwischen den Skalen entlang des Postulats der Alternativhypothese. Mit anderen Worten: Die Probanden geben auf der Objektiv-Skala ein signifikant geringeres Generalisiertes Vertrauen an, als auf der Standard-Skala. Hierbei ist allerdings einzuschränken, dass es sich bei den Resultaten für die Kontraste PR_1 vs. PT_4 und PR_2 vs. PT_3 aufgrund der geringen Teststärke auch um falsch-signifikante Ergebnisse handeln könnte.

Tab. 7.61: Ergebnisse der parametrischen Signifikanzprüfung (T-Test für unabhängige Stichproben)

Between-subject-Design (Standard-Skala vs. Objektiv-Skala)	
PR ₁ vs. PR ₂	T(84,299)=-3,413 (einseitig); p<.01; g=,698; 1-β=,868
PR ₁ vs. PT ₄	T(86,991)=-2,311 (einseitig); p<.05; g=,467; 1-β=,757
PR ₂ vs. PT ₃	T(91,705)=2,274 (einseitig); p<.05; g=,456; 1-β=,374

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

Zwischen den Messergebnissen der Objektiv-Kurzskala und der Standard-Skala ergeben sich nur für die Kontraste PR_{1(KS)} vs. PR₂ und PR₂ vs. PT_{3(KS)} überzufällige Unterschiede. Dabei kann das signifikante Ergebnis für keine der Paarungen statistisch abgesichert werden. In beiden Fällen liegt die post-hoc-Teststärke unterhalb des per Konvention festgelegten Schwellenwertes. Die Wahrscheinlichkeit, eine richtige Alternativhypothese anzunehmen bzw. die Nullhypothese abzulehnen, gilt damit als vermindert. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Der Unterschied zwischen PR_{1(KS)} und PT₄ wird hingegen nicht signifikant. Es liegt somit keine überzufällig große Differenz zwischen den Messergebnissen der Skalen vor. Dabei ergibt sich allerdings eine sehr geringe Teststärke, aus der eine stark erhöhte β-Fehler-Wahrscheinlichkeit resultiert. Die Wahrscheinlichkeit, die Nullhypothese abzulehnen, bzw. die Wahrscheinlichkeit, ein gültige Alternativhypothese anzunehmen, sind demnach stark reduziert, während gleichzeitig das Risiko, eine korrekte Alternativhypothese abzulehnen, stark gesteigert ist. Offensichtlich scheint die Datengrundlage zu gering zu sein, als dass hier eine eindeutige Entscheidung zu Gunsten der H₁ getroffen werden konnte. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Ausgehend von der Effektstärke ergeben sich zwischen den Skalen ausschließlich kleine Effekte. Die Differenzen der Messergebnisse sind demnach zwar sehr gering, allerdings noch praktisch bedeutsam und von null verschieden. Somit scheint auch zwischen PR_{1(KS)} und PT₄ eine relevante Differenz zu bestehen, die jedoch nicht überzufällig ausfällt. Diese Interpretation wird für die Skalenpaarungen PR_{1(KS)} und PR₂ sowie PR_{1(KS)} und PT₄ durch den Wertebereich des jeweiligen 95%-Konfidenzintervall der Differenz bestätigt. Der Wert 0 wird hier jeweils nicht überstrichen. Für den Kontrast von PR₂ vs. PT_{3(KS)} enthält das 95%-Konfidenzintervall der Differenz dagegen auch den Wert 0, sodass die Möglichkeit besteht, dass der wahre Wert für die Abweichung auch einer Nulldifferenz entsprechen könnte. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

Entsprechend des signifikanten Ergebnisses wird für den Kontrast von PR_{1(KS)} vs. PR₂ und PR₂ vs. PT_{3(KS)} einstweilen die H₁ beibehalten und die H₀ abgelehnt. Für PR_{1(KS)} vs. PT₄ wird dagegen zunächst die H₀ beibehalten und die H₁ abgelehnt. Aufgrund der geringen Teststärken der einzelnen Signifikanzprüfungen besteht allerdings die Gefahr von falsch-signifikanten bzw.

7. Hauptstudie

falsch-nicht-signifikanten Ergebnissen. Unabhängig von der Überzufälligkeit der Differenzen zwischen den Vertrauensskalen, kann für jede der Paarungen festgestellt werden, dass die Probanden auf der Objektiv-Kurzskala einen niedrigeren Vertrauenswert erreichen als auf der Standard-Skala.

Tab. 7.62: Ergebnisse der parametrischen Signifikanzprüfung

Between-subject-Design (Standard-Skala vs. Objektiv-Kurzskala)	
PR_{1(KS)} vs. PR₂	T(93,513)=-2,351 (einseitig); p<.05; g=.473; 1-β=.759
PR_{1(KS)} vs. PT₄	T(96,040)=-1,327 (zweiseitig); n.s.; g=.264; 1-β=.262
PR₂ vs. PT_{3(KS)}	T(90,029)=1,888 (einseitig); p<.05; g=.379; 1-β=.601

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow$ *kleiner Effekt*, $d = .50 \rightarrow$ *mittlerer Effekt*, $d = .80 \rightarrow$ *großer Effekt* (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

7.2.6.7.2.2 Nonparametrische Signifikanzprüfung

Entsprechend der Empfehlungen von BORTZ/LIENERT/BOEHNKE sollen im nachfolgenden Abschnitt die Ergebnisse der parametrischen Verfahren anhand von verteilungsfreien Verfahren validiert werden (Bortz/Lienert/Boehnke (2008), S. 81, S. 85f).

Bei den in die Signifikanzprüfung eingehenden Versuchsgruppen handelt es sich um wechselseitig unabhängige Stichproben, weshalb zur Ermittlung des Unterschieds zwischen der Objektiv- und der Standard-Skala der Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben zum Einsatz kommt. Es ergeben sich dann die folgenden Hypothesen:

H₀: Die Mediane der Objektiv-Skala bzw. -Kurzskala zur Erfassung Generalisierten Vertrauens (PR₁, PT₃) unterscheiden sich nicht vom Median der Standard-Skala zur Erfassung Generalisierten Vertrauens (PR₂, PT₄). In der Grundgesamtheit liegen gleich viele Rangplatzüberschreitungen wie -unterschreitungen vor.

$$\tilde{x}_{PR1} = \tilde{x}_{PR2}$$

$$\tilde{x}_{PR} = \tilde{x}_{PT}$$

$$\tilde{x}_{PT} = \tilde{x}_{PR2}$$

H₁: Die Mediane der Objektiv-Skala bzw. -Kurzskala zur Erfassung Generalisierten Vertrauens (PR₁, PT₃) sind kleiner als der Median der Standard-Skala zur Erfassung Generalisierten Vertrauens (PR₂, PT₄). In der Grundgesamtheit liegen mehr Rangplatzunterschreitungen als Rangplatzüberschreitungen vor.

$$\tilde{x}_{PR1} < \tilde{x}_{PR2}$$

$$\tilde{x}_{PR} < \tilde{x}_{PT4}$$

$$\tilde{x}_{PT} < \tilde{x}_{PR}$$

Die Signifikanzprüfung der Differenz zwischen den Messergebnissen von Objektiv- und Standard-Skala ergab für jede der Skalenpaarungen einen überzufälligen Unterschied. Dabei konnte allerdings keines der signifikanten Ergebnisse statistisch abgesichert werden. Vielmehr

7. Hauptstudie

liegt die Teststärke der betrachteten Kontraste jeweils weit unterhalb des Schwellenwertes von $1-\beta=.80$. Die Wahrscheinlichkeit, eine korrekte Alternativhypothese anzunehmen, gilt damit als verringert. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärke zeigt je nach betrachteter Paarungen einen schwachen bis mittleren Effekt an ($.10 \leq \omega \leq .50$). Es scheint sich demnach gemessen an den Normwerten für Effektstärken um geringe, allerdings praktisch relevante, Differenzen zwischen den Messergebnissen von Objektiv- und Standard-Skala zu handeln. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 668f; Bühner/Ziegler (2009), S. 177ff, S. 269f, S. 633; Cohen (1988), S. 9f, S. 223ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Betrachtet man die mittleren Ränge, so ergeben sich für die Versuchspersonen, deren Generalisiertes Vertrauen mit Hilfe der Objektiv-Skala erfasst wurde, im Durchschnitt niedrigere Ränge als für die Versuchspersonen, deren Generalisiertes Vertrauen anhand der Standard-Skala erfasst wurde. Die Richtung des Unterschieds zwischen beiden Messverfahren entspricht somit dem von der Alternativhypothese vorausgesagten Trend. (Bühner/Ziegler (2009), S. 287)

In allen drei Fällen wird die H_0 abgelehnt und die H_1 beibehalten. Der Unterschied zwischen den Messmethoden entspricht demnach der Vorhersage der Alternativhypothese. Die Probanden erreichen somit auf der Objektiv-Skala überzufällig niedrigere Vertrauenswerte als auf der Standard-Skala. Aufgrund der geringen post-hoc-Teststärke besteht allerdings das Risiko eines falsch-signifikanten Ergebnisses.

Tab. 7.63: Ergebnisse der nonparametrischen Signifikanzprüfung (Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)

Between-subject-Design (Standard- vs. Objektiv-Skala)	
PR₁ vs. PR₂	$z=-3,220$ (einseitig); $p<.01$; $\varphi=.322$; $1-\beta=.291$
PR₁ vs. PT₄	$z=-2,232$ (einseitig); $p<.01$; $\varphi=.221$; $1-\beta=.105$
PR₂ vs. PT₃	$z=-2,179$ (einseitig); $p<.05$; $\varphi=.216$; $1-\beta=.279$

z : empirischer z -Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p>.05$ (zweiseitig))

* $p<.05$ (einseitig), ** $p<.01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

Für den Unterschied zwischen Objektiv-Kurzskala und Standard-Skala ergeben sich nur bei der Gegenüberstellung von PR_{1(KS)} und PR₂ sowie PR₂ und PT_{3(KS)} signifikante Ergebnisse. Beide können allerdings nicht statistisch abgesichert werden, da die post-hoc-Teststärke zu niedrig ausfällt. Die Wahrscheinlichkeit, dass durch den statistischen Test der postulierte Populationseffekt aufgedeckt wird, sofern er wahr ist, ist somit reduziert. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Für den Kontrast von PR_{1(KS)} vs. PT₄ ergibt sich dagegen kein signifikanter Unterschied, wobei aber auch dieses Resultat nicht statistisch abgesichert werden kann. Dabei geht die niedrige Teststärke mit einer erhöhten β -Fehler-Wahrscheinlichkeit einher. Die Chancen, dass eine gültige Alternativhypothese angenommen wird, sind somit stark reduziert und das Risiko einer fälschlichen Entscheidung zugunsten der H_0 stark erhöht. Vor dem Hintergrund, dass sich kein signifikantes Ergebnis eingestellt hat, ist somit keine eindeutige Interpretation möglich.

7. Hauptstudie

Höchstwahrscheinlich ist die Datenbasis, auf der der durchgeführte Signifikanztest fußt, zu gering, als dass eine Entscheidung zugunsten der Alternativhypothese fallen konnte. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärke ist bei allen drei Paarungen im Bereich für schwache Effekte angesiedelt ($.10 \leq \omega \leq .30$). Zwischen den Messergebnissen von Objektiv-Kurzskala und Standard-Skala scheinen demnach lediglich geringe Abweichungen zu existieren, die allerdings trotzdem als praktisch relevant eingestuft werden können. Insbesondere für die Skalenpaarung $PR_{1(KS)}$ und PT_4 kann die von null verschiedene Effektstärke als Hinweis auf ein falsch-nicht-signifikantes Ergebnis gewertet werden.

Bei der Betrachtung der mittleren Ränge ergibt sich das bereits bei den vorangegangenen Kontrastierungen von Objektiv- und Standard-Messung beobachtete Verhältnis. Versuchspersonen, die Objektiv-Kurzskala bearbeitet hatten, erreichen im Mittel niedrigere Ränge als Versuchspersonen, die die Standard-Skala bearbeitet hatten. Die Richtung des Unterschieds im Antwortverhalten auf den beiden Skalen scheint somit der Behauptung der Alternativhypothese zu entsprechen. (Bühner/Ziegler (2009), S. 287)

Ausgehend von den Resultaten der Signifikanzprüfungen werden für die Kontraste $PR_{1(KS)}$ vs. PR_2 und PR_2 vs. $PT_{3(KS)}$ die H_1 beibehalten und die H_0 abgelehnt. Dabei bleibt zu bedenken, dass es sich aufgrund der jeweils geringen Teststärke auch um falsch-signifikante Ergebnisse handeln könnte. Bezüglich des Unterschiedes von $PR_{1(KS)}$ und PT_4 wird dagegen die H_0 beibehalten und die H_1 abgelehnt. Hierbei ist allerdings die Gefahr eines β -Fehlers zu bedenken.

Tab. 7.64: Ergebnisse der nonparametrischen Signifikanzprüfung (Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)

Between-subject-Design (Standard-Skala vs. Objektiv-Kurzskala)	
$PR_{1(KS)}$ vs. PR_2	$z = -2,229$ (einseitig); $p < .05$; $\varphi = .223$; $1-\beta = .287$
$PR_{1(KS)}$ vs. PT_4	$z = -1,340$ (zweiseitig); n.s.; $\varphi = .133$; $1-\beta = .100$
PR_2 vs. $PT_{3(KS)}$	$z = -1,827$ (einseitig); $p < .05$; $\varphi = .181$; $1-\beta = .224$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f;

Cohen (1988), S. 223ff)

1- β : Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

7.2.6.7.2.3 Zusammenfassung

Auch im Rahmen der als *between-subject*-Design angeordneten Versuchsgruppen ergeben sich zwischen den Objektiv-Skalen und der Standard-Skala ausschließlich durchschnittliche Differenzen, deren Betrag unterhalb eines halben Skalenpunktes rangiert. Der Kontrast zwischen den miteinander verglichenen Vertrauensskalen fällt demnach vergleichsweise schwach aus. Hieraus lässt sich ableiten, dass es im Messergebnis strenggenommen wohl kaum einen Unterschied macht, ob zur Erfassung Generalisierten Vertrauens die Objektiv-Skalen oder die Standard-Skala herangezogen werden. Was die Überzufälligkeit der gefundenen Differenzen angeht, so ergeben sich je nach Skalenpaarung widerstreitende Ergebnisse.

Auch im Rahmen des *between-subject*-Designs ergeben sich zwischen den Messergebnissen der Objektiv-Kurzskala und der Standard-Skala geringere Abweichungen als zwischen der

Objektiv-Skala und der Standard-Skala. Analog zum gleichen Befund beim *within-subject*-Design scheint die Objektiv-Kurzskala damit eine größere inhaltliche Ähnlichkeit zur Standard-Skala aufzuweisen als die längere Objektiv-Skala.

Zunächst ist festzustellen, dass aus den Resultaten der parametrischen und nonparametrischen Verfahren eine wechselseitige Bestätigung der signifikanten bzw. nicht-signifikanten Ergebnisse hervorgeht. Entsprechend der von BORTZ/LIENERT/BOEHNKE vorgeschlagenen Parallelprüfung wird daher angenommen, dass die Resultate jeweils Gültigkeit aufweisen. Über alle Skalenvergleiche hinweg betrachtet, wurde dabei in der Mehrheit der Fälle ein überzufälliger Unterschied zwischen dem Antwortverhalten auf der Objektiv-Skala bzw. der Objektiv-Kurzskala und dem Antwortverhalten auf der Standard-Skala festgestellt. Unabhängig von der Signifikanz des Ergebnisses machten die Probanden auf den Objektiv-Skalen verglichen mit der Standard-Skala geringere Angaben hinsichtlich ihrer Ausstattung mit Generalisiertem Vertrauen.

Eine mögliche Interpretation der signifikanten Ergebnisse ist die Annahme, dass die Probanden entweder auf den Objektiv-Skalen oder auf der Standard-Skala keine wahrheitsgemäßen Angaben hinsichtlich ihres Vertrauenslevels gemacht haben. Ob sie dabei über ein in Wahrheit geringer ausgeprägtes Generalisiertes Vertrauen als auf der Standard-Skala angegeben verfügen oder ob sie in Wahrheit über ein stärker ausgeprägtes Generalisiertes Vertrauen als auf der Objektiv-Skala angegeben verfügen, kann an dieser Stelle nicht zweifelsfrei festgestellt werden.

Die nicht-signifikanten Ergebnisse zeigen indes an, dass es sich bei den Objektiv-Skalen und der Standard-Skala um mindestens äquivalente Messverfahren handelt. Aufgrund der zweiseitigen Signifikanzprüfung kann hier davon ausgegangen werden, dass sich die Messergebnisse der verschiedenen Vertrauensskalen wechselseitig entsprechen und weder überzufällige Abweichungen nach unten noch nach oben auftreten.

Zu den Ergebnissen der Signifikanzprüfungen im *between-subject*-Design ist zu bemerken, dass hier jeweils unterschiedlich zusammengesetzte Probandengruppen gegeneinander getestet wurden. Effekte auf die Messergebnisse, die sich aus der Charakterdispositionen der Versuchspersonen ableiten und den Unterschied zwischen den Messergebnissen beeinflussen könnten, konnten demnach nicht mit absoluter Sicherheit konstant gehalten werden. Es ist daher nicht vollkommen auszuschließen, dass neben der Art und Weise der Messung auch andere Aspekte Auswirkungen auf die gefundenen Unterschiede zwischen den Messverfahren haben. Allerdings kann, zumindest was das Geschlechterverhältnis und die Altersstruktur angeht, für die personelle Zusammensetzung der Versuchsgruppen weitestgehend wechselseitige Äquivalenz angenommen werden.

7.2.6.8 Effekt des Treatments (Identifikation der Messintention)

Im folgenden Abschnitt soll der Unterschied zwischen dem Antwortverhalten auf den Objektiv-Skalen und dem Antwortverhalten auf der Standard-Skala differenziert nach der Identifikation der Messintention der Objektiv-Skalen betrachtet werden.

Die Anordnung der Versuchsgruppen im *within-subject*-Design bzw. *between-subject*-Design hat hierbei besondere Implikationen für die in die jeweiligen Analysen eingehenden Fallzahlen. So kann eine Differenzierung der Probanden anhand der Identifikation der Messintention nur aus den Objektiv-Messungen abgeleitet werden. Im Zuge von Standard-Messungen wurde generell keine gesonderte Abfrage der Messintention getätigt, da hier bereits Augenscheinvalidität gegeben war.

7. Hauptstudie

Für die als *within-subject*-Design konzipierten Versuchsgruppen hat dies lediglich zur Konsequenz, dass in die Unterschiedsprüfung von Objektiv-Skalen und Standard-Skala je nachdem, ob die Zieldimension der Objektiv-Skalen aufgedeckt wurde oder nicht, unterschiedliche Fallzahlen eingehen. Mit anderen Worten: Die Probandengruppe, die die Messintention erkannt hat, fällt u.U. größer oder kleiner aus als die Probandengruppe, die die Messintention nicht erkannt hat oder umgekehrt. Dies ist damit zu begründen, dass es sich jeweils um verbundene Stichproben handelt, die anhand der Information aus der jeweiligen Objektiv-Messung in die verschiedenen Probandengruppen unterschieden werden.

Für die Versuchsgruppen, die als *between-subject*-Design angeordnet werden können, ergeben sich dagegen schwerwiegendere Auswirkungen. Die Tatsache, dass es sich um unverbundene Stichproben handelt, hat zur Folge, dass innerhalb jeder Kontrastierung von Objektiv-Skalen und Standard-Skala eine Asymmetrie der Fallzahlen vorliegt. Dies ergibt sich daraus, dass die Information hinsichtlich der Identifikation der Messintention an die jeweils eingehende Objektiv-Skala bzw. -Kurzskala gebunden ist und keine Auswirkung auf die jeweils Eingang findende Standard-Skala hat. Mit anderen Worten: Es wird im Prinzip nur auf Seiten der Objektiv-Skalen nach der Identifikation der Zieldimension differenziert, während auf Seiten der Standard-Skala jeweils alle Probanden, die eine Antwort abgegeben haben, in die Unterschiedsprüfung eingehen. Die Konsequenz hieraus können je nachdem, ob Varianzunterschiede zwischen den unterschiedlich großen kontrastierten Stichproben existieren, eher progressive oder eher konservative Testentscheidungen⁸⁶ sein.

Im Rahmen des *within-subject*-Design werden sowohl für die Probandengruppe, die die Messintention der Objektiv-Skalen erkannt haben als auch für die Probandengruppe, die die Messintention der Objektiv-Skalen nicht erkannt haben, die folgenden Kontraste geprüft:

$$PR_1 \text{ vs. } PT_1$$

$$PR_{1(KS)} \text{ vs. } PT_1$$

$$PR_2 \text{ vs. } PT_2$$

$$PR_2 \text{ vs. } PT_{2(KS)}$$

Im Rahmen des *between-subject*-Design werden unter den Bedingungen, dass die Zieldimension der Objektiv-Skalen identifiziert wurde bzw. nicht identifiziert wurde, jeweils die folgenden Unterschiedsprüfungen durchgeführt:

$$PT_3 \text{ vs. } PT_4$$

$$PT_{3(KS)} \text{ vs. } PT_4$$

$$PR_1 \text{ vs. } PR_2$$

$$PR_{1(KS)} \text{ vs. } PR_2$$

$$PR_1 \text{ vs. } PT_4$$

$$PR_{1(KS)} \text{ vs. } PT_4$$

$$PR_2 \text{ vs. } PT_3$$

$$PR_2 \text{ vs. } PT_{3(KS)}$$

⁸⁶ „Sind die Stichprobenumfänge deutlich unterschieden, wird die Präzision des t-Tests nicht beeinträchtigt, solange die Varianzen gleich sind. Sind jedoch weder die Stichprobenumfänge noch die Varianzen gleich, ist mit einem erheblich höheren Prozentsatz an Fehlentscheidungen zu rechnen. Nach Ramsey (1980) entscheidet ein Test eher zu Gunsten der H_1 , wenn die Varianz in der kleineren Stichprobe größer ist als die Varianz in der größeren Stichprobe (progressive Testentscheidung). Ist die Varianz in der größeren Stichprobe jedoch größer als in der kleineren, fallen die Testentscheidungen eher konservativ, d.h., zugunsten der H_0 , aus. Insbesondere progressive Fehlentscheidungen sind zu vermeiden, da dabei mit einer erhöhten Wahrscheinlichkeit auf Unterschiede geschlossen werden kann, die faktisch nicht vorhanden sind.“ (Bortz (2005), S. 141)

Wie bereits bei den zuvor beschriebenen Kontrasten besteht das übergeordnete Erkenntnisziel dieser Analyse darin, herauszufinden, ob sich das Antwortverhalten auf den Objektiv-Skalen von dem auf der Standard-Skala unterscheidet. Im Speziellen soll nun untersucht werden, ob der in den vorherigen Signifikanzprüfungen gefundene Unterschied weiterhin Bestand hat, wenn man die Probanden nach der Identifikation der Messintention der Objektiv-Skalen differenziert. Ergeben sich dabei nur für die Probandengruppe, die die Messintention nicht erkannt hat, überzufällige Unterschiede in den Messergebnissen, stellt dies einen Hinweis darauf dar, dass die erfolgreiche Verschleierung der Messintention einen Beitrag zur Differenz zwischen den Messresultaten leistet. Mit anderen Worten: Die Verwendung Objektiver Tests zur Erfassung Generalisierten Vertrauens wirkt sich auf das Antwortverhalten der Probanden aus.

Ergeben sich dagegen in keiner der Probandengruppen nach der Differenzierung noch Unterschiede oder ergeben sich in beiden Probandengruppen Unterschiede, so stellt dies einen Hinweis darauf dar, dass irgendein anderes Phänomen für die Differenz verantwortlich sein muss.

7.2.6.8.1 *Within-Subject-Design*

Bei den im Rahmen des *within-subject*-Design in die Signifikanzprüfung eingehenden Fälle der verschiedenen Messverfahrenspaarungen handelt es sich jeweils um verbundene Stichproben. Da die Abfrage der Identifikation der Messintention ausschließlich im Nachhinein der Objektiv-Skalen erfolgt, werden so auch auf Seiten der Standard-Skala die Fälle entlang dieser Variable differenziert. Auf diese Weise wird jeweils das Antwortverhalten der gleichen Fälle auf den verschiedenen Vertrauensskalen verglichen. Bei der Interpretation der nachfolgenden Ergebnisse ist nun allerdings zu beachten, dass die Zieldimension der Standard-Skala für alle Probanden grundsätzlich einsehbar war und somit strenggenommen die Anzahl der Fälle, der die Messintention dieser Skala bewusst war, prinzipiell unterschätzt wird.

7.2.6.8.1.1 Parametrische Signifikanzprüfung

Zunächst sollen die Ergebnisse der parametrischen Signifikanzprüfung betrachtet werden. Zum Einsatz kam hierbei der T-Test für abhängige Stichproben.

Es ergeben sich die folgenden Hypothesen, die sowohl für die Probandengruppe, die die Messintention der Objektiv-Skalen erkannt hat (Gruppe A), als auch für die Probandengruppe, die die Messintention der Objektiv-Skalen nicht erkannt hat (Gruppe B), Gültigkeit besitzen:

H₀: Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR₁ bzw. PT₂) zur Erfassung Generalisierten Vertrauens und der Standard-Skala (PT₁ bzw. PR₂) zur Erfassung Generalisierten Vertrauens fallen innerhalb der Population gleich aus bzw. die Differenz der Messwertpaare in der Grundgesamtheit ist gleich null. Abweichungen zwischen den Mittelwerten von Objektiv-Skala bzw. -Kurzskala und Standard-Skala sind rein zufällig.

$$\mu_{PR1} = \mu_{PT1} \text{ bzw. } \mu_d = 0$$

$$\mu_{PR2} = \mu_{PT2} \text{ bzw. } \mu_d = 0$$

H₁: Der Mittelwert der Objektiv-Skala bzw. -Kurzskala (PR₁) zur Erfassung Generalisierten Vertrauens fällt innerhalb der Population jeweils niedriger aus als der Mittelwert der

Standard-Skala (PT₁) zur Erfassung Generalisierten Vertrauens bzw. die Differenz der Messwertpaare innerhalb der Grundgesamtheit ist kleiner null.

$$\mu_{PR1} < \mu_{PT1} \text{ oder } \mu_d < 0$$

bzw.

Der Mittelwert der Standard-Skala (PR₂) zur Erfassung Generalisierten Vertrauens fällt innerhalb der Population höher aus als der jeweilige Mittelwert der Objektiv-Skala bzw. -Kurzskala (PT₂) zur Erfassung Generalisierten Vertrauens bzw. die Differenz der Messwertpaare innerhalb der Grundgesamtheit ist größer null.

$$\mu_{PR2} > \mu_{PT2} \text{ oder } \mu_d > 0$$

Da die in Gruppe A zusammengefassten Probanden die Zieldimension der Objektiv-Skalen aufdecken konnten, wird angenommen, dass sie zu einer willentlichen Abstimmung ihrer Antworten auf den Objektiv-Skalen mit ihren Antworten auf der Standard-Skala befähigt sind. Innerhalb von Gruppe A sollte sich demnach eher keine überzufällige Differenz zwischen den Skalenwerten der Objektiv-Skalen und der Standard-Skala ergeben. In Gruppe B wurde die Maskierung der Messintention hingegen nicht aufgedeckt, sodass hier mit einem signifikanten Unterschied zwischen den Skalenwerten der Objektiv-Skalen und der Standard-Skala gerechnet wird. Ausgehend von diesen Überlegungen ist für Gruppe A der Eintritt des Postulats der H₀ mit einer höheren Wahrscheinlichkeit zu erwarten, wohingegen für Gruppe B eine höhere Eintrittswahrscheinlichkeit der H₁ erwartet wird.

Für Gruppe A ergeben sich in *Versuchsgruppe 1* und *Versuchsgruppe 2* sowohl zwischen der Objektiv-Skala und der Standard-Skala als auch zwischen der Objektiv-Kurzskala und der Standard-Skala überzufällige Unterschiede im Antwortverhalten. Die signifikanten bzw. hochsignifikanten Ergebnisse können dabei allerdings in keinem der Fälle statistisch abgesichert werden. Die post-hoc-Teststärke rangiert jeweils unterhalb des Schwellenwertes von $1-\beta=.80$. Das bedeutet: Die Wahrscheinlichkeit, zu Gunsten einer wahren Alternativhypothese zu entscheiden, ist reduziert. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärke zeigt je nach Skalenpaarung einen mittleren bis großen Effekt an. Trotz der eher geringen durchschnittlichen Differenz zwischen den Messwerten der Skalen, scheint es sich demnach zunächst um praktisch bedeutsame Differenzen im Antwortverhalten zu handeln. Diese Interpretation wird allerdings nur im Falle des Kontrastes von PR₂ vs. PT₂ durch das 95%-Konfidenzintervall der Differenz unterstützt. In den übrigen Fällen beinhaltet der jeweilige Vertrauensbereich den Wert 0, sodass auch eine Nulldifferenz zwischen den Messergebnissen der Skalen wahrscheinlich ist. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

In allen vier Fällen wird zunächst die H₀ abgelehnt und die H₁ beibehalten. Der Unterschied im Antwortverhalten zwischen den Skalen scheint entlang des Postulats der Alternativhypothese zu verlaufen. Die Probanden geben in Gruppe A somit auf den Objektiv-Skalen ein signifikant geringeres Generalisiertes Vertrauen an als auf der Standard-Skala. Hierbei ist allerdings einzuschränken, dass es sich bei den signifikanten Resultaten aufgrund der geringen Teststärke auch um falsch-signifikante Ergebnisse handeln könnte.

Tab. 7.65: Ergebnisse der parametrischen Signifikanzprüfung (T-Test für abhängige Stichproben)

Within-subject-Design [Messintention der Objektiv-Skalen identifiziert]	
PR₁ vs. PT₁	T(11)= -2,159 (einseitig); p<,05; g=,623; 1-β=,647
PR_{1(KS)} vs. PT₁	T(11)= -1,915 (einseitig); p<,05; g=,553; 1-β=,560
PR₂ vs. PT₂	T(23)= 2,882 (einseitig); p<,01; g=,588; 1-β=,649
PR₂ vs. PT_{2(KS)}	T(23)= 2,015 (einseitig); p<,05; g=,411; 1-β=,621

T(df): empirischem t-Wert inklusive Freiheitsgrade
(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow \text{kleiner Effekt}$, $d = .50 \rightarrow \text{mittlerer Effekt}$, $d = .80 \rightarrow \text{großer Effekt}$ (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 247

(Quelle: eigene Berechnungen)

Für Gruppe B ergeben sich für alle Skalenpaarungen, außer für den Kontrast von PR₂ vs. PT_{2(KS)}, hochsignifikante Unterschiede im Antwortverhalten. Dabei kann allerdings ausschließlich das Ergebnis der Signifikanzprüfung von PR₁ vs. PT₁ statistisch abgesichert werden. Für die übrigen Kontraste erreicht die Teststärke jeweils nicht die Schwelle von 1-β=.80, was in einer Verminderung der Wahrscheinlichkeit, eine korrekte Alternativhypothese anzunehmen, resultiert. Im Falle von PR₂ vs. PT_{2(KS)} erhöht sich dadurch die β-Fehler-Wahrscheinlichkeit auf einen Wert weit oberhalb des akzeptierten Niveaus von β=.20. Die Wahrscheinlichkeit, die Nullhypothese abzulehnen, bzw. die Wahrscheinlichkeit, eine gültige Alternativhypothese anzunehmen, sind hier stark reduziert, während gleichzeitig das Risiko, eine korrekte Alternativhypothese abzulehnen, stark gesteigert ist. Offensichtlich scheint die Datengrundlage zu gering zu sein, als dass hier eine eindeutige Entscheidung zu Gunsten der H₁ getroffen werden konnte. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärken bewegen sich im Bereich schwacher und moderater Effekte ($.20 \leq g \leq .80$). Es scheint sich demnach um praktisch bedeutsame Unterschiede im Antwortverhalten zwischen den Skalen zu handeln. Dieser Eindruck bestätigt sich bei Betrachtung des 95%-Konfidenzintervall der Differenz für die Kontraste mit hochsignifikanten Ergebnissen. Für die Skalenpaarung PR₂ vs. PT_{2(KS)} enthält das zugehörige Konfidenzintervall allerdings den Wert 0, sodass hier mit einer hohen Wahrscheinlichkeit in Wahrheit keine Differenz zwischen den Messergebnissen existiert. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

Für die Kontraste PR₁ vs. PT₁, PR_{1(KS)} vs. PT₁ und PR₂ vs. PT₂ werden jeweils die H₀ abgelehnt und die H₁ beibehalten. Für die genannten Paarungen wird somit innerhalb von Gruppe B angenommen, dass sich die Probanden auf den Objektiv-Skalen weniger vertrauensvoll darstellen als auf der Standard-Skala. Für PR_{1(KS)} vs. PT₁ und PR₂ vs. PT₂ ist allerdings einschränkend zu bemerken, dass es sich aufgrund der geringen Teststärke auch um ein falsch-signifikantes Ergebnis handeln könnte.

Im Falle von PR₂ vs. PT_{2(KS)} wird dagegen die H₀ beibehalten und die H₁ abgelehnt. Zwar unterscheiden sich die Messergebnisse entsprechend der durch die Alternativhypothese vorgegebenen Richtung, allerdings scheint die Differenz nicht überzufällig groß zu sein. Aufgrund

7. Hauptstudie

des nicht signifikanten Ergebnisses infolge einer zweiseitigen Signifikanzprüfung können Objektiv-Kurzskala und Standard-Skala in diesem Fall als äquivalente Messverfahren betrachtet werden. Beide Skalen unterscheiden sich weder in Richtung der H_1 noch in deren Gegenrichtung. Vielmehr weisen die Messergebnisse keinen Unterschied auf. Vor dem Hintergrund der geringen Teststärke und der erhöhten β -Fehler-Wahrscheinlichkeit kann allerdings nicht ausgeschlossen werden, dass hier ein Fehler zweiter Art vorliegt bzw. das Ergebnis in Wahrheit falsch-nicht-signifikant ist.

Tab. 7.66: Ergebnisse der parametrischen Signifikanzprüfung (T-Test für abhängige Stichproben)

Within-subject-Design [Messintention der Objektiv-Skalen <u>nicht</u> identifiziert]	
PR₁ vs. PT₁	T(37)= -4,093 (einseitig); p<,01; g=,664; 1- β =,947
PR_{1(KS)} vs. PT₁	T(37)= -2,920 (einseitig); p<,01; g=,474; 1- β =,687
PR₂ vs. PT₂	T(25)= 2,540 (einseitig); p<,01; g=,498; 1- β =,530
PR₂ vs. PT_{2(KS)}	T(25)= 1,162 (zweiseitig); n.s.; g=,228; 1- β =,201

T(df): empirischem t-Wert inklusive Freiheitsgrade
(einseitig) bzw. (zweiseitig): Testrichtung
* p<.05 (einseitig), ** p<.01 (einseitig)
n.s.: nicht signifikant (p>.05 (zweiseitig))
g: Effektstärke mit $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff; Bühner/Ziegler (2009), S. 187, S. 251f)
1- β : Teststärke
Vgl. Bühner/Ziegler (2009), S. 247
(Quelle: eigene Berechnungen)

7.2.6.8.1.2 Nonparametrische Signifikanzprüfung

Nachfolgend sollen die Ergebnisse der parametrischen Signifikanzprüfung anhand von verteilungsfreien Verfahren validiert werden. Zum Einsatz kommt dabei der Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben. Entsprechend ergeben sich für die Probandengruppe, welche die Messintention der Objektiv-Skalen erkannt hat (Gruppe A), und die Probandengruppe, welche die Messintention der Objektiv-Skalen nicht erkannt hat (Gruppe B), prinzipiell die folgenden Hypothesen:

H_0 : Die Mediane der Objektiv-Skala bzw. -Kurzskala (PR₁ bzw. PT₂) zur Erfassung generalisierten und der Standard-Skala (PT₁ bzw. PR₂) zur Erfassung Generalisierten Vertrauens fallen in der Grundgesamtheit gleich aus bzw. die Rangsummen der beiden Messvarianten sind in der Population gleich. Unterschiede zwischen der Objektiv-Skala bzw. -Kurzskala und der Standard-Skala sind rein zufällig.

$$W^+ = W^- \text{ bzw. } \tilde{x}_{PR} = \tilde{x}_{PT1} \text{ bzw. } \tilde{x}_{PR2} = \tilde{x}_{PT2}$$

H_1 : Der Median der Objektiv-Skala bzw. -Kurzskala (PR₁) fällt innerhalb der Grundgesamtheit niedriger aus als der Median der Standard-Messung (PT₁). Zwischen den beiden Messvarianten treten positive Ränge mit größerer Häufigkeit auf als negative Ränge.

$$W^+ > W^- \text{ bzw. } \tilde{x}_{PR1} < \tilde{x}_{PT1}$$

bzw.

Der Median der Standard-Messung (PR₂) fällt innerhalb der Grundgesamtheit höher aus als der Median der Objektiv-Skala bzw. -Kurzskala (PT₂). Zwischen den beiden Messvarianten treten negative Ränge mit größerer Häufigkeit auf als positive Ränge.

$$W^+ < W^- \text{ bzw. } \tilde{x}_{PR2} > \tilde{x}_{PT}$$

Auch hier ist zu beachten, dass aufgrund der Tatsache, dass in Gruppe A die Messintention der Objektiv-Skalen identifiziert wurde und damit im Prinzip die Möglichkeit zu einer bewussten Angleichung des Antwortverhaltens auf den verschiedenen Vertrauensskalen gegeben ist, kein überzufälliger Unterschied zwischen den Skalenwerten der Objektiv-Skalen und der Standard-Skala erwartet wird. Für Gruppe B wird dagegen erwartet, dass aufgrund der erfolgreichen Verschleierung der Messintention der Objektiv-Skalen eine überzufällige Differenz zwischen den Skalenwerten der Objektiv-Skalen und der Standard-Skala existiert. Mit anderen Worten: Innerhalb von Gruppe A sollte sich eher die H_0 als gültig erweisen, während sich für Gruppe B eine Tendenz zur H_1 ergeben sollte.

In Gruppe A ergibt sich ausschließlich für die Paarung PR_2 vs. PT_2 ein überzufälliger Unterschied zwischen dem Antwortverhalten auf den verschiedenen Vertrauensskalen, wobei das hochsignifikante Ergebnis nicht statistisch abgesichert werden kann. Die übrigen Kontraste werden nicht signifikant und auch diese Ergebnisse können nicht statistisch abgesichert werden. In allen vier Fällen liegt damit die post-hoc-Teststärke unterhalb des Schwellenwertes von $1-\beta=.80$. Mit Ausnahme von PR_2 vs. PT_2 geht dieser Umstand Hand in Hand mit einer erhöhten β -Fehler-Wahrscheinlichkeit. (Bortz (2005), S. 122; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärken der Skalenpaarungen nehmen Werte an, die die Differenzen als mittlere bis starke Effekte ausweisen ($\omega \geq .30$). Der Unterschied zwischen den Messergebnissen von Objektiv- und Standard-Skala bzw. Objektiv-Kurzskala und Standard-Skala scheint demnach praktisch relevant zu sein. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 270, S. 633; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Für die aus *Versuchsgruppe 1* stammenden Kontraste (PR_1 vs. PT_1 und $PR_{1(KS)}$ vs. PT_1) ergeben sich jeweils mehr Fälle, für die das per Standard-Skala ermittelte Generalisierte Vertrauen höher ausfällt als das per Objektiv-Skala ermittelte Generalisierte Vertrauen. Entsprechend der zugehörigen Alternativhypothese fällt die Rangsumme für positive Ränge höher aus als die Rangsumme für negative Ränge.

Für die aus *Versuchsgruppe 2* stammenden Kontraste (PR_2 vs. PT_2 und $PR_{2(KS)}$ vs. PT_2) ergeben sich jeweils mehr Fälle, für die das per Objektiv-Skala ermittelte Generalisierte Vertrauen niedriger ausfällt als das per Standard-Skala ermittelte Generalisierte Vertrauen. Entsprechend der zugehörigen Alternativhypothese fällt die Rangsumme für negative Ränge höher aus als die Rangsumme für positive Ränge.

Ausgehend von der z-verteilten Prüfgröße ergibt sich ausschließlich aus der Differenz der Rangsummen von PR_2 vs. PT_2 ein hochsignifikantes Ergebnis. Es wird dementsprechend die H_1 beibehalten und die H_0 abgelehnt. Die Probanden geben hier auf der Objektiv-Skala somit einen niedrigeren Vertrauenslevel an als auf der Standard-Skala. Die übrigen Kontraste entsprechen zwar hinsichtlich der Richtung der Differenz dem Postulat der Alternativhypothese, allerdings resultiert kein überzufälliger Unterschied. Es werden demnach jeweils die H_1 abgelehnt und die H_0 beibehalten. Ausgehend von der zweiseitigen Signifikanzprüfung weicht damit das Messergebnis der Objektiv-Skalen weder nach unten noch nach oben vom Messergebnis der Standard-Skala ab und umgekehrt. Objektiv-Skalen und Standard-Skalen scheinen hier somit äquivalente Messverfahren darzustellen.

Einschränkend ist zu bemerken, dass die niedrige Teststärke und bei den nicht-signifikanten Ergebnissen die resultierende erhöhte β -Fehler-Wahrscheinlichkeit auch falsch-signifikante bzw. falsch-nicht signifikante Ergebnisse bewirken kann.

**Tab. 7.67: Ergebnisse der nonparametrischen Signifikanzprüfung
(Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben)**

Within-subject-Design [Messintention der Objektiv-Skalen identifiziert]	
PR₁ vs. PT₁	$z = -1,890$ (zweiseitig); n.s.; $\varphi = ,546$; $1-\beta = ,390$
PR_{1(KS)} vs. PT₁	$z = -1,732$ (zweiseitig); n.s.; $\varphi = ,500$; $1-\beta = ,337$
PR₂ vs. PT₂	$z = -2,517$ (einseitig); $p < ,01$; $\varphi = ,514$; $1-\beta = ,493$
PR₂ vs. PT_{2(KS)}	$z = -1,897$ (zweiseitig); n.s.; $\varphi = ,387$; $1-\beta = ,425$

z : empirischer z -Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > ,05$ (zweiseitig))

* $p < ,05$ (einseitig), ** $p < ,01$ (einseitig)

φ : Effektstärke mit $\omega = ,10$ (kleiner Effekt), $\omega = ,30$ (moderater Effekt), $\omega = ,50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 278

(Quelle: eigene Berechnungen)

In Gruppe B zeigen sich für alle Paarungen, außer für PR₂ vs. PT_{2(KS)}, überzufällige Unterschiede, die allerdings nur für den Vergleich von PR₁ und PT₁, d.h. für den Vergleich der Objektiv-Skala mit der Standard-Skala aus *Versuchsgruppe 1*, statistisch abgesichert werden können. Das bedeutet: Nur für diese Fälle ist die Wahrscheinlichkeit, eine wahre Alternativhypothese anzunehmen, ausreichend hoch. Für den Kontrast PR₂ vs. PT_{2(KS)} ergibt sich darüber hinaus eine β -Fehler-Wahrscheinlichkeit oberhalb der per Konvention festgelegten Schwelle von $\beta = ,20$. Die Wahrscheinlichkeit, eine korrekte Alternativhypothese irrtümlich abzulehnen, ist damit stark erhöht. (Bortz (2005), S. 122; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärken nehmen Werte im Spektrum für schwache bis starke Effekte an ($\omega \geq ,10$). Die Differenzen zwischen Objektiv-Skalen und Standard-Skala können demnach als praktisch bedeutsam angesehen werden. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 270, S. 633; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Für die aus *Versuchsgruppe 1* stammenden Kontraste (PR₁ vs. PT₁ und PR_{1(KS)} vs. PT₁) ergeben sich jeweils mehr Fälle, für die das per Standard-Skala ermittelte Generalisierte Vertrauen höher ausfällt als das per Objektiv-Skala ermittelte Generalisierte Vertrauen. Entsprechend der zugehörigen Alternativhypothese fällt die Rangsumme für positive Ränge höher aus als die Rangsumme für negative Ränge.

Für die aus *Versuchsgruppe 2* stammenden Kontraste (PR₂ vs. PT₂ und PR_{2(KS)} vs. PT₂) ergeben sich jeweils mehr Fälle, für die das per Objektiv-Skala ermittelte Generalisierte Vertrauen niedriger ausfällt als das per Standard-Skala ermittelte Generalisierte Vertrauen. Entsprechend der zugehörigen Alternativhypothese fällt die Rangsumme für negative Ränge höher aus als die Rangsumme für positive Ränge.

Der z -verteilten Prüfgröße entsprechend wird für die Kontraste PR₁ vs. PT₁, PR_{1(KS)} vs. PT₁ und PR₂ vs. PT₂ die H_1 beibehalten und die H_0 abgelehnt. Die Probanden haben bei diesen Skalenpaarungen auf der jeweiligen Objektiv-Skala bzw. -Kurzskala ein signifikant niedrigeres Generalisiertes Vertrauen angegeben als auf der Standard-Skala, was der Vorhersage der Alternativhypothese entspricht.

Auch die Differenz zwischen PR₂ und PT_{2(KS)} entspricht der Richtung der Alternativhypothese, allerdings wird der Unterschied nicht signifikant. Hier wird demnach die H_1 abgelehnt und die H_0 beibehalten. Die durchgeführte zweiseitige Signifikanzprüfung weist außerdem darauf hin, dass es hier zwischen Objektiv-Kurzskala und Standard-Skala zu gar keiner Abweichung kommt. Die

7. Hauptstudie

Ergebnisse der beiden Messverfahren bzw. die Messverfahren selbst können als äquivalent angesehen werden.

Aufgrund der niedrigen Teststärke bei der Gegenüberstellung der Objektiv-Kurzskala und der Standard-Skala in den *Versuchsgruppen 1* und *2* und der erhöhten β -Fehler-Wahrscheinlichkeit für PR_2 vs. $PT_{2(KS)}$ sind falsch signifikante bzw. falsch nicht-signifikante Ergebnisse nicht auszuschließen.

**Tab. 7.68: Ergebnisse der nonparametrischen Signifikanzprüfung
(Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben)**

Within-subject-Design [Messintention der Objektiv-Skalen nicht identifiziert]	
PR_1 vs. PT_1	$z=-3,441$ (einseitig); $p<,01$; $\varphi=,558$; $1-\beta=,817$
$PR_{1(KS)}$ vs. PT_1	$z=-2,668$ (einseitig); $p<,01$; $\varphi=,433$; $1-\beta=,572$
PR_2 vs. PT_2	$z=-2,309$ (einseitig); $p<,05$; $\varphi=,453$; $1-\beta=,708$
PR_2 vs. $PT_{2(KS)}$	$z=-1,155$ (zweiseitig); n.s.; $\varphi=,227$; $1-\beta=,192$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p>,05$ (zweiseitig))

* $p<,05$ (einseitig), ** $p<,01$ (einseitig)

φ : Effektstärke mit $\omega = ,10$ (kleiner Effekt), $\omega = ,30$ (moderater Effekt), $\omega = ,50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 278

(Quelle: eigene Berechnungen)

7.2.6.8.2 Between-subject-Design

Bei den im Rahmen des *between-subject*-Design in die Signifikanzprüfung eingehenden Fälle der verschiedenen Messverfahrenspaarungen handelt es sich jeweils um unverbundene Stichproben. Es wird somit jeweils das Antwortverhalten unterschiedlicher Probanden miteinander verglichen. Die Differenzierung nach der Identifikation der Messintention erfolgt dabei ausschließlich auf Seiten der Objektiv-Skalen, da die Zieldimension der Standard-Skala jeweils offen lag und davon ausgegangen werden konnte, dass diese dadurch allen Probanden bekannt war. Auf diese Weise gehen für die Standard-Skala grundsätzlich alle Fälle, die diese bearbeitet haben, und damit prinzipiell mehr Fälle als für die Objektiv-Skalen in die jeweiligen Unterschiedsprüfungen ein. Diese Asymmetrie der Fallzahlen kann einen progressiven oder konservativen Effekt auf die Testentscheidungen haben.

7.2.6.8.2.1 Parametrische Signifikanzprüfung

Zunächst sollen die Ergebnisse der parametrischen Signifikanzprüfung betrachtet werden. Hierbei wurde der T-Test für unabhängige Stichproben eingesetzt.

Zu prüfen sind die folgenden Hypothesen, die sowohl für die Probandengruppe, die die Messintention der Objektiv-Skalen erkannt hat (Gruppe A), als auch für die Probandengruppe, die die Messintention der Objektiv-Skalen nicht erkannt hat (Gruppe B), getestet werden sollen:

H_0 : Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR_1 , PT_3) zur Erfassung Generalisierten Vertrauens unterscheiden sich jeweils nicht vom Mittelwert der Standard-Skala (PR_2 , PT_4) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwert ergibt null.

$$\mu_{PT3} = \mu_{PT4} \text{ oder } \mu_{PT3} - \mu_{PT4} = 0$$

$$\mu_{PR1} = \mu_{PR2} \text{ oder } \mu_{PR1} - \mu_{PR2} = 0$$

7. Hauptstudie

$$\mu_{PR1} = \mu_{PT4} \text{ oder } \mu_{PR1} - \mu_{PT4} = 0$$

$$\mu_{PR2} = \mu_{PT3} \text{ oder } \mu_{PR2} - \mu_{PT3} = 0$$

H₁: Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR₁, PT₃) zur Erfassung Generalisierten Vertrauens fallen jeweils niedriger aus als der Mittelwert der Standard-Skala (PR₂, PT₄) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwerte ist kleiner als null.

$$\mu_{PT3} < \mu_{PT4} \text{ oder } \mu_{PT3} - \mu_{PT4} < 0$$

$$\mu_{PR1} < \mu_{PR2} \text{ oder } \mu_{PR1} - \mu_{PR2} < 0$$

$$\mu_{PR1} < \mu_{PT4} \text{ oder } \mu_{PR1} - \mu_{PT4} < 0$$

bzw.

Der Mittelwert der Standard-Skala (PR₂, PT₄) zur Erfassung Generalisierten Vertrauens fällt jeweils höher aus als die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR₁, PT₃) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwerte ist größer als null.

$$\mu_{PR2} > \mu_{PT3} \text{ oder } \mu_{PR2} - \mu_{PT3} > 0$$

Im Rahmen der Signifikanzprüfung ist zu beachten, dass für Gruppe A eine höhere Eintrittswahrscheinlichkeit für die H₀ angenommen wird, während für Gruppe B ein Trend zur H₁ erwartet wird.

Tab. 7.69: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Between-subject-Design [Messintention der Objektiv-Skalen identifiziert]	
PT ₃ vs. PT ₄	T(61)= ,306 (zweiseitig); n.s.; g=,102; 1-β=,061
PR ₁ vs. PR ₂	T(60)= -2,067 (einseitig); p<,05; g=,664; 1-β=,654
PR ₁ vs. PT ₄	T(22,493)= -1,781 (einseitig); p<,05; g=,460; 1-β=,411
PR ₂ vs. PT ₃	T(59)= -,261 (zweiseitig); n.s.; g=,087; 1-β=,058

T(df): empirischem t-Wert inklusive Freiheitsgrade
(einseitig) bzw. (zweiseitig): Testrichtung
*p<,05 (einseitig), **p<,01 (einseitig)
n.s.: nicht signifikant (p>,05 (zweiseitig))
g: Effektstärke mit d = .20 → *kleiner Effekt*, d = .50 → *mittlerer Effekt*, d = .80 → *großer Effekt* (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)
1-β: Teststärke
Vgl. Bühner/Ziegler (2009), S. 260
(Quelle: eigene Berechnungen)

Tab. 7.70: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Between-subject-Design (Kurzskala) [Messintention der Objektiv-Skalen identifiziert]	
PT _{3(KS)} vs. PT ₄	T(30,303)= -,063 (zweiseitig); n.s.; g=,014; 1-β=,050
PR _{1(KS)} vs. PR ₂	T(60)= -1,308 (zweiseitig); n.s.; g=,421; 1-β=,252
PR _{1(KS)} vs. PT ₄	T(62)= -,729 (zweiseitig); n.s.; g=,233; 1-β=,111
PR ₂ vs. PT _{3(KS)}	T(29,980)= -,734 (zweiseitig); n.s.; g=,212; 1-β=,096

T(df): empirischem t-Wert inklusive Freiheitsgrade
(einseitig) bzw. (zweiseitig): Testrichtung
*p<,05 (einseitig), **p<,01 (einseitig)
n.s.: nicht signifikant (p>,05 (zweiseitig))
g: Effektstärke mit d = .20 → *kleiner Effekt*, d = .50 → *mittlerer Effekt*, d = .80 → *großer Effekt* (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)
1-β: Teststärke
Vgl. Bühner/Ziegler (2009), S. 260
(Quelle: eigene Berechnungen)

7. Hauptstudie

Vergleicht man die Mittelwerte der verschiedenen Messvarianten miteinander, so zeigt sich, dass für alle Kontrastierungen von Objektiv-Skala bzw. -Kurzskala und Standard-Skala, mit Ausnahme von PT_3 vs. PT_4 , die Probanden auf der Objektiv-Skala im Durchschnitt ein geringeres Generalisiertes Vertrauen angeben als die Probanden, die die Standard-Skala bearbeitet hatten. In der Mehrzahl der Fälle entspricht die Relation der Skalenwerte somit dem Postulat der H_1 . Für den Vergleich von PT_3 und PT_4 ergibt sich dagegen die umgekehrte Relation.

Für die Gegenüberstellung der Objektiv-Skalen und der Standard-Skala innerhalb von Gruppe A ergaben sich nur für die Kontraste PR_1 vs. PR_2 und PR_1 vs. PT_4 signifikante Unterschiede. Unabhängig davon konnte keines der Ergebnisse der Signifikanzprüfungen statistisch abgesichert werden bzw. die post-hoc-Teststärke liegt durchgängig unterhalb des notwendigen Niveaus von $1-\beta=.80$. Die Wahrscheinlichkeit, zu Gunsten der H_1 zu entscheiden, sofern diese tatsächlich Gültigkeit besitzt, ist somit reduziert. Gleichzeitig liegt die β -Fehler-Wahrscheinlichkeit im Falle der nicht-signifikanten Ergebnisse jeweils weit oberhalb des zulässigen Grenzwertes von $\beta=.20$, sodass hier das Risiko, eine korrekte Alternativhypothese abzulehnen, stark gesteigert ist. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

In diesem Zusammenhang ist außerdem zu beachten, dass sich für die Gegenüberstellung von PT_1 und PT_4 , $PT_{3(KS)}$ und PT_4 sowie PR_2 und $PT_{3(KS)}$ signifikante Varianzunterschiede zwischen den eingehenden Stichproben ergeben, die so gelagert sind, dass konservative Testentscheidungen begünstigt werden. Mit anderen Worten: Die Varianz in der größeren Stichprobe (hier: Standard-Messung) fällt jeweils größer aus als in der kleineren Stichprobe (hier: Objektiv-Messung). Für PR_1 vs. PT_4 scheint diese Tendenz aufgrund des überzufälligen Unterschiedes zwischen den beiden Messvarianten keine Relevanz zu besitzen. Für die beiden übrigen Kontraste erhöht sich im Zusammenspiel mit der jeweils hohen β -Fehler-Wahrscheinlichkeit, das Risiko eines falsch-nicht-signifikanten Resultates. (Bortz (2005), S. 141)

Die Effektstärken bewegen sich je nach Paarung in einem Bereich unterhalb der Schwelle für einen schwachen Unterschied bis hin zu Werten, die einem moderaten Unterschied ($.20 < g \leq .80$) zwischen den Messvarianten entsprechen. Es scheinen sich demnach nur teilweise praktisch bedeutsame Unterschiede im Antwortverhalten zwischen den Skalen einzustellen. Allerdings enthält in allen Fällen, mit Ausnahme von PR_1 vs. PR_2 und $PR_{1(KS)}$ vs. PR_2 , das 95%-Konfidenzintervall der Differenz den Wert null, sodass mit einer hohen Wahrscheinlichkeit angenommen werden kann, dass innerhalb weiter Teile von Gruppe A kein Unterschied zwischen den Mittelwerten der verschiedenen Messmethoden besteht. Für den Kontrast PR_1 vs. PR_2 deutet der Befund, dass das 95%-Konfidenzintervall der Differenz den Wert null nicht enthält, daraufhin, dass hier tatsächlich ein überzufälliger Unterschied zwischen den Messvarianten vorliegt. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

Für die Kontraste PR_1 vs. PR_2 und PR_1 vs. PT_4 wird aufgrund der im Vergleich hohen Effekt- und Teststärke einstweilen die H_1 beibehalten und die H_0 verworfen. Es wird für diese Paarungen somit angenommen, dass sich die Probanden auf der Objektiv-Skala weniger vertrauensvoll darstellen als auf der Standard-Skala. Zu PR_1 vs. PT_4 ist außerdem anzumerken, dass sich trotz der Richtung des ermittelten Varianzunterschiedes zwischen den in die Signifikanzprüfung eingehenden Stichproben, die eine Tendenz zu einer konservativen Testentscheidung impliziert, ein signifikanter Unterschied ergeben hat.

7. Hauptstudie

Für die verbliebenen Kontraste wird jeweils die H_0 beibehalten und die H_1 abgelehnt. Es wird daher angenommen, dass sich das Antwortverhalten der Probanden auf den verschiedenen Messskalen zur Erfassung Generalisierten Vertrauens innerhalb von Gruppe A mehrheitlich nicht unterscheidet. Die Objektiv-Skalen und die Standard-Skala scheinen somit unter der Bedingung, dass die Messintention der Objektiv-Skalen identifiziert wurde, äquivalente Messverfahren darzustellen.

Tab. 7.71: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Between-subject-Design [Messintention der Objektiv-Skalen <u>nicht</u> identifiziert]	
PT₃ vs. PT₄	T(88,991)= -1,634 (zweiseitig); n.s.; $g=.332$; $1-\beta=.341$
PR₁ vs. PR₂	T(83,712)= -3,184 (einseitig); $p<.01$; $g=.646$; $1-\beta=.735$
PR₁ vs. PT₄	T(86,068)= -2,122 (einseitig); $p<.05$; $g=.422$; $1-\beta=.624$
PR₂ vs. PT₃	T(86,942)= -2,631 (einseitig); $p<.01$; $g=.543$; $1-\beta=.570$

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

* $p<.05$ (einseitig), ** $p<.01$ (einseitig)

n.s.: nicht signifikant ($p>.05$ (zweiseitig))

g : Effektstärke mit $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

$1-\beta$: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

Tab. 7.72: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Between-subject-Design (Kurzskala) [Messintention der Objektiv-Skalen <u>nicht</u> identifiziert]	
PT_{3(KS)} vs. PT₄	T(88,715)= -.918 (zweiseitig); n.s.; $g=.188$; $1-\beta=.142$
PR_{1(KS)} vs. PR₂	T(85,704)= -1,906 (einseitig); $p<.05$; $g=.480$; $1-\beta=.715$
PR_{1(KS)} vs. PT₄	T(87,836)= -1,315 (zweiseitig); n.s.; $g=.266$; $1-\beta=.234$
PR₂ vs. PT_{3(KS)}	T(86,855)= -1,885 (einseitig); $p<.05$; $g=.392$; $1-\beta=.570$

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

* $p<.05$ (einseitig), ** $p<.01$ (einseitig)

n.s.: nicht signifikant ($p>.05$ (zweiseitig))

g : Effektstärke mit $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

$1-\beta$: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

Die Betrachtung der Mittelwerte ergibt, dass deren Relation für jede der innerhalb von Gruppe B vorgenommenen Kontrastierungen dem Postulat der H_1 entsprechen.

Innerhalb von Gruppe B ergeben sich für alle Paarungen, außer für PT₃ vs. PT₄, PT_{3(KS)} vs. PT₄ und PR_{1(KS)} vs. PT₄, signifikante und teilweise auch hochsignifikante Unterschiede zwischen den durchschnittlichen Summenwerten der Objektiv-Skalen und der Standard-Skala. Unabhängig davon kann allerdings keines der Resultate statistisch abgesichert werden. In allen Fällen liegt die post-hoc-Teststärke unterhalb des per Konvention festgelegten Schwellenwertes ($1-\beta<.80$). Für die Paarungen, für die sich kein überzufälliger Kontrast zwischen den Messvarianten Generalisierten Vertrauens ergibt, hat die niedrige Teststärke eine β -Fehler-Wahrscheinlichkeit oberhalb des Grenzwertes von $\beta=.20$ zur Folge. Insgesamt bedeutet dies, dass die Wahrscheinlichkeit, eine richtige Alternativhypothese anzunehmen bzw. eine falsche Nullhypothese abzulehnen, als vermindert angesehen werden kann. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die in die jeweiligen Signifikanzprüfungen eingehenden Stichproben weisen signifikante Varianzunterschiede auf. Dies deutet daraufhin, dass auf den Objektiv-Skalen tatsächlich anders

geantwortet wurde als auf der Standard-Skala. Die Richtung der Varianzunterschiede impliziert dabei eine Tendenz zu konservativen Testentscheidungen, d.h. den Vorzug der H_0 . So fällt die Varianz der größeren Stichprobe (hier: Standard-Messung) grundsätzlich höher aus als die Varianz in der kleineren Stichprobe (hier: Objektiv-Messung). Vor diesem Hintergrund sind die nicht-signifikanten Unterschiede kritisch zu hinterfragen bzw. es erhöht sich für die Paarungen die Gefahr von falsch-nicht-signifikanten Ergebnissen. (Bortz (2005), S. 141)

Die Werte der Effektstärke zeigen in der Mehrzahl der Fälle kleine bis mittlere Effekte ($.20 \leq g \leq .80$) an. Die Effektstärke der Paarung $PT_{3(KS)}$ vs. PT_4 liegt dabei nur knapp unterhalb des Schwellenwertes für einen kleinen Effekt. Es scheinen somit praktisch bedeutsame Unterschiede zwischen den verschiedenen Messvarianten vorzuliegen. Für alle nicht-signifikanten Kontraste enthält das 95%-Konfidenzintervall der Differenz jeweils den Wert null, sodass für diese Konstellationen auch eine Nulldifferenz zwischen den Skalenmittelwerten im Bereich des Möglichen liegt. Bei den Paarungen, die signifikante Unterschiede im Antwortverhalten auf den verschiedenen Vertrauensskalen aufweisen, überstreicht das 95%-Konfidenzintervall der Differenz jeweils nicht den Wert null. Eine Ausnahme bildet hierbei PR_2 vs. $PT_{3(KS)}$. In der Mehrzahl dieser Fälle kann somit mit hoher Wahrscheinlichkeit angenommen werden, dass tatsächlich ein von null verschiedener Unterschied zwischen den Messverfahren existiert. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

Vorläufig wird für die Gegenüberstellung von PT_3 und PT_4 , $PT_{3(KS)}$ und PT_4 sowie $PR_{1(KS)}$ und PT_4 die H_0 beibehalten und die H_1 abgelehnt, auch wenn für diese Paarungen eine deutliche Affinität zu konservativen Testentscheidungen festgestellt wurde. Ebenso fallen die Teststärken vergleichsweise niedrig aus, was ein Ansteigen der β -Fehler-Wahrscheinlichkeit zur Folge hat. Falsch-nicht-signifikante Ergebnisse können hier somit nicht ausgeschlossen werden. Allerdings kann angenommen werden, dass es sich bei den Objektiv-Skalen und der Standard-Skala mindestens um zueinander äquivalente Messverfahren handelt.

Für die übrigen Paarungen wird die H_1 beibehalten und die H_0 abgelehnt. Für die Mehrzahl der Kontraste wird somit angenommen, dass sich die Probanden, wenn sie die Messintention der Objektiv-Skala nicht identifizieren konnten, auf der Standard-Skala vertrauensvoller darstellen als auf den Objektiv-Skalen.

7.2.6.8.2.2 Nonparametrische Signifikanzprüfung

Zum Zweck der Validierung der Ergebnisse der parametrischen Signifikanzprüfungen werden nachfolgend die gleichen Kontraste mit Hilfe nonparametrischer Verfahren untersucht. Hierzu wird der Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben eingesetzt.

Die innerhalb der Probandengruppe, welche die Messintention der Objektiv-Skalen erkannt hat (Gruppe A), und der Probandengruppe, welche die entsprechende Messintention nicht erkannt hat (Gruppe B), zu prüfenden Kontraste lassen sich in den folgenden Hypothesen zusammenfassen:

H_0 : Die Mediane der Objektiv-Skala bzw. -Kurzskala zur Erfassung Generalisierten Vertrauens (PR_1 , PT_3) unterscheiden sich nicht vom Median der Standard-Skala zur Erfassung Generalisierten Vertrauens (PR_2 , PT_4). In der Grundgesamtheit liegen gleich viele Rangplatzüberschreitungen wie -unterschreitungen vor.

7. Hauptstudie

$$\tilde{x}_{PT} = \tilde{x}_{PT4}$$

$$\tilde{x}_{PR1} = \tilde{x}_{PR}$$

$$\tilde{x}_{PR1} = \tilde{x}_{PT4}$$

$$\tilde{x}_{PT3} = \tilde{x}_{PR}$$

H₁: Die Mediane der Objektiv-Skala bzw. -Kurzskala zur Erfassung Generalisierten Vertrauens (PR₁, PT₃) sind kleiner als der Median der Standard-Skala zur Erfassung Generalisierten Vertrauens (PR₂, PT₄). In der Grundgesamtheit liegen mehr Rangplatzunterschreitungen als Rangplatzüberschreitungen vor.

$$\tilde{x}_{PT} < \tilde{x}_{PT4}$$

$$\tilde{x}_{PR1} < \tilde{x}_{PR2}$$

$$\tilde{x}_{PR1} < \tilde{x}_{PT4}$$

$$\tilde{x}_{PT} < \tilde{x}_{PR}$$

Auch für die Signifikanzprüfung mit Hilfe der nonparametrischen Verfahren wird angenommen, dass die H₀ in Gruppe A eine höhere Eintrittswahrscheinlichkeit gegenüber der H₁ besitzt, während für Gruppe B das Gegenteil, d.h. eine höhere Eintrittswahrscheinlichkeit der H₁ gegenüber der H₀, angenommen wird.

**Tab. 7.73: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design [Messintention der Objektiv-Skalen identifiziert]	
PT ₃ vs. PT ₄	z=-,252 (zweiseitig); n.s.; φ = ,032; 1- β = ,051
PR ₁ vs. PR ₂	z=-2,073 (einseitig); p<,05; φ = ,263; 1- β = ,196
PR ₁ vs. PT ₄	z=-1,488 (zweiseitig); n.s.; φ = ,186; 1- β = ,086
PR ₂ vs. PT ₃	z=-,233 (zweiseitig); n.s.; φ = ,030; 1- β = ,051

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant (p>,05 (zweiseitig))

*p<,05 (einseitig), **p<,01 (einseitig)

φ : Effektstärke mit ω = ,10 (kleiner Effekt), ω = ,30 (moderater Effekt), ω = ,50 (starker Effekt) (Bühner/Ziegler (2009), S. 269f;

Cohen (1988), S. 223ff)

1- β : Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

**Tab. 7.74: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design (Kurzskala) [Messintention der Objektiv-Skalen identifiziert]	
PT _{3(KS)} vs. PT ₄	z=-,205 (zweiseitig); n.s.; φ = ,026; 1- β = ,051
PR _{1(KS)} vs. PR ₂	z=-1,180 (zweiseitig); n.s.; φ = ,150; 1- β = ,073
PR _{1(KS)} vs. PT ₄	z=-,695 (zweiseitig); n.s.; φ = ,087; 1- β = ,058
PR ₂ vs. PT _{3(KS)}	z=-,761 (zweiseitig); n.s.; φ = ,097; 1- β = ,059

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant (p>,05 (zweiseitig))

*p<,05 (einseitig), **p<,01 (einseitig)

φ : Effektstärke mit ω = ,10 (kleiner Effekt), ω = ,30 (moderater Effekt), ω = ,50 (starker Effekt) (Bühner/Ziegler (2009), S. 269f;

Cohen (1988), S. 223ff)

1- β : Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

Die Signifikanzprüfung der Differenz zwischen den Messergebnissen von Objektiv-Skalen und der Standard-Skala ergab innerhalb von Gruppe A für keine der Skalenpaarungen, mit Ausnahme des Kontrastes PR₁ vs. PR₂, einen überzufälligen Unterschied. Allerdings konnte keines der Ergebnisse statistisch abgesichert werden. Die post-hoc-Teststärke lag für jede der Paarungen

7. Hauptstudie

weit unterhalb des Grenzwertes ($1-\beta < .80$). Für die Gegenüberstellung von PR_1 und PR_2 ergibt sich dadurch die Gefahr eines Alpha-Fehlers. Für die übrigen Kontraste geht die niedrige Teststärke mit einer erhöhten β -Fehler-Wahrscheinlichkeit einher. Das bedeutet: Es könnte sich um falsch-nicht-signifikante Ergebnisse handeln. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärke nimmt je nach betrachteter Paarung Werte an, die keinen Effekt oder maximal einen kleinen Effekt anzeigen ($\omega < .30$). Die Differenzen zwischen den Messmethoden scheinen demnach kaum praktische Relevanz zu besitzen. Am größten fällt der Unterschied noch für den signifikanten Kontrast von PR_1 und PR_2 aus. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 668f; Bühner/Ziegler (2009), S. 177ff, S. 269f, S. 633; Cohen (1988), S. 9f, S. 223ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Betrachtet man die mittleren Ränge, so ergeben sich in der Mehrzahl der Paarungen für die Probanden, deren Generalisiertes Vertrauen mit Hilfe der Objektiv-Skala erfasst wurde, im Durchschnitt niedrigere Ränge als für die Probanden, deren Generalisiertes Vertrauen anhand der Standard-Skala erfasst wurde. Die Richtung des Unterschieds zwischen beiden Messverfahren entspricht somit dem von der Alternativhypothese vorausgesagten Trend. Eine Ausnahme hiervon bildet die Kontrastierung von PT_3 vs. PT_4 . Hier verläuft die Relation umgekehrt, was darauf hindeutet, dass die Probanden sich auf der Objektiv-Skala vertrauensvoller darstellen als auf der Standard-Skala. (Bühner/Ziegler (2009), S. 287)

Mit Ausnahme von PR_1 vs. PR_2 wird für alle Signifikanzprüfungen die H_0 beibehalten und die H_1 abgelehnt. Es wird daher angenommen, dass es sich, insofern die Probanden die Zieldimension der Objektiv-Skalen identifizieren können, bei den Objektiv-Skalen und der Standard-Skala um äquivalente Messverfahren zur Erfassung Generalisierten Vertrauens handelt.

**Tab. 7.75: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design [Messintention der Objektiv-Skalen <u>nicht</u> identifiziert]	
PT_3 vs. PT_4	$z = -1,585$ (zweiseitig); n.s.; $\varphi = ,166$; $1-\beta = ,118$
PR_1 vs. PR_2	$z = -2,898$ (einseitig); $p < .01$; $\varphi = ,309$; $1-\beta = ,172$
PR_1 vs. PT_4	$z = -1,980$ (einseitig); $p < .05$; $\varphi = ,209$; $1-\beta = ,243$
PR_2 vs. PT_3	$z = -2,444$ (einseitig); $p < .01$; $\varphi = ,259$; $1-\beta = ,123$

z : empirischer z -Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f;

Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

**Tab. 7.76: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design (Kurzska) [Messintention der Objektiv-Skalen nicht identifiziert]	
PT _{3(KS)} vs. PT ₄	$z = -.938$ (zweiseitig); n.s.; $\varphi = .098$; $1-\beta = .073$
PR _{1(KS)} vs. PR ₂	$z = -2.175$ (einseitig); $p < .05$; $\varphi = .232$; $1-\beta = .274$
PR _{1(KS)} vs. PT ₄	$z = -1.308$ (zweiseitig); n.s.; $\varphi = .138$; $1-\beta = .096$
PR ₂ vs. PT _{3(KS)}	$z = -1.756$ (einseitig); $p < .05$; $\varphi = .186$; $1-\beta = .212$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f;

Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

Innerhalb von Gruppe B ergeben sich mehrheitlich signifikante Unterschiede im Antwortverhalten auf den Objektiv-Skalen und der Standard-Skala. Hiervon weichen nur die Kontraste PT₃ vs. PT₄, PT_{3(KS)} vs. PT₄ und PR_{1(KS)} vs. PT₄ ab. Allerdings kann keines der Ergebnisse statistisch abgesichert werden, da die post-hoc-Teststärken jeweils unterhalb des Schwellenwertes von $1-\beta = .80$ liegen. Gleichzeitig ergibt sich für die nicht-signifikanten Kontraste eine gesteigerte β -Fehler-Wahrscheinlichkeit ($\beta > .20$). Die Wahrscheinlichkeit zu Gunsten der H_1 zu entscheiden, wenn die H_1 tatsächlich gilt, ist somit reduziert und das Ergebnis ist nicht eindeutig interpretierbar. Möglicherweise handelt es sich um falsch-nicht-signifikante Ergebnisse. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärke zeigt je nach betrachteter Paarung keinen bis einen mittleren Effekt an ($.10 < \omega < .50$). Es scheint sich demnach gemessen an den Normwerten für Effektstärken um in der Mehrzahl geringe, allerdings praktisch relevante, Differenzen zwischen den Messergebnissen der Objektiv-Skalen und der Standard-Skala zu handeln. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 668f; Bühner/Ziegler (2009), S. 177ff, S. 269f, S. 633; Cohen (1988), S. 9f, S. 223ff; Cohen (1992), S. 156, Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Aus den Werten für die mittleren Ränge geht hervor, dass die Probanden, die die Objektiv-Skalen bearbeitet haben, im Durchschnitt niedrigere Ränge erreichen, als die Probanden, die die Standard-Skala bearbeitet haben. Dies stimmt mit der durch die H_1 postulierten Relation überein. (Bühner/Ziegler (2009), S. 287)

Für die Mehrzahl der Paarungen innerhalb von Gruppe B wird die H_1 beibehalten und die H_0 verworfen. Ausgenommen hiervon sind lediglich PT₃ vs. PT₄, PT_{3(KS)} vs. PT₄ und PR_{1(KS)} vs. PT₄, für die vorläufig die H_0 beibehalten und die H_1 abgelehnt wird. Entsprechend der überzufälligen Unterschiede im Antwortverhalten bedeutet dies, dass insofern die Probanden die Messintention der Objektiv-Skalen nicht identifizieren können, sich diese auf der Standard-Skala tendenziell vertrauensvoller darstellen als auf den Objektiv-Skalen. Aus den nicht-signifikanten Ergebnissen lässt sich zusätzlich ablesen, dass die Objektiv-Skalen mindestens äquivalente Messmethoden zur Erfassung Generalisierten Vertrauens darstellen.

7.2.6.8.3 Zusammenfassung

Im vorangegangenen Abschnitt wurden die durchschnittlichen Summenwerte der Objektiv-Skalen mit den durchschnittlichen Summenwerten der Standard-Skala unter der Bedingung kontrastiert, dass entweder die Zieldimension der Objektiv-Skalen identifiziert oder nicht identifiziert wurde. Es wurde dabei angenommen, dass sich im Falle der Identifikation der Messintention die Skalensummenwerte eher angleichen sollten, während die Nicht-Identifikation der Messintention eine überzufällige Differenz der Skalensummenwerte herbeiführt. Diese Differenz sollte dabei dergestalt gelagert sein, dass die Probanden sich auf den Objektiv-Skalen im Durchschnitt weniger vertrauensvoll präsentieren als auf der Standard-Skala.

Zunächst ist festzustellen, dass die durchschnittliche Differenz zwischen den Summenscores der Objektiv-Skalen und der Standard-Skala weder in der Probandengruppe, die die Messintention der Objektiv-Skalen identifiziert hat (Gruppe A), noch in der Probandengruppe, die die Messintention nicht identifiziert hat (Gruppe B), und das sowohl im Rahmen des *within-subject*-Design als auch im Rahmen des *between-subject*-Design, den Wert eines halben Skalenpunktes überschreitet. Der Unterschied zwischen den Messergebnissen ist somit als marginal zu bezeichnen. Ausgehend hiervon scheint es damit strenggenommen kaum einen Unterschied zu machen, ob zur Erfassung Generalisierten Vertrauens die im Rahmen der Pilotstudie entwickelten Objektiv-Skalen verwendet werden oder die seit vielen Jahren standardmäßig verwendete *Trust in People-Scale* (Standard-Skala) zum Einsatz kommt. Unabhängig vom Betrag der Differenz zwischen den Messergebnissen, erweist sich der Unterschied im Antwortverhalten jedoch innerhalb von Gruppe B (keine Identifikation der Messintention) eher als praktisch bedeutsam als in Gruppe A (Identifikation der Messintention).

Betrachtet man die Resultate der Signifikanzprüfungen innerhalb des *within-subject*-Designs, ergibt sich, dass die Ergebnisse der parametrischen und nonparametrischen Verfahren für Gruppe B deckungsgleich ausfallen. In Gruppe B sind dabei, mit Ausnahme von PR_2 vs. $PT_{2(KS)}$, überzufällige Abweichungen im Antwortverhalten zwischen den Objektiv-Skalen und der Standard-Skala festzustellen. Hierbei haben die Probanden auf den Objektiv-Skalen ein signifikant niedrigeres Generalisiertes Vertrauen angegeben als auf der Standard-Skala. Für den Kontrast von PR_2 und $PT_{2(KS)}$ gilt indes, dass Standard-Skala und Objektiv-Kurzskala offensichtlich äquivalente Messmethoden darstellen.

Dagegen lassen sich innerhalb von Gruppe A, die für die für die Kontraste PR_1 vs. PT_1 , $PR_{1(KS)}$ vs. PT_1 und PR_2 vs. $PT_{2(KS)}$ im Rahmen der parametrischen Verfahren gefundenen signifikanten Ergebnisse durch die nonparametrischen Verfahren nicht replizieren. Es ergeben sich somit widerstreitende Ergebnisse. Ausgehend von den niedrigen Teststärken wird daher angenommen, dass sich das Antwortverhalten zwischen den Objektiv- und Standard-Skalen in Gruppe A nicht unterscheidet. Die Messverfahren können daher als äquivalent angenommen werden. Anders verhält es sich bei PR_2 vs. PT_2 , hier sind die Ergebnisse von parametrischen und nonparametrischen Verfahren deckungsgleich und der Unterschied im Antwortverhalten wird gleichermaßen signifikant. Die Probanden stellen sich hier auf der Objektiv-Skala weniger vertrauensvoll dar als auf der Standard-Skala.

Für den signifikanten Unterschied zur Standard-Skala im Rahmen der parametrischen Signifikanzprüfung könnten hier auch inhaltliche Besonderheiten der Objektiv-Skalen verantwortlich sein. Im Vergleich zur Standard-Skala enthalten die Objektiv-Skalen eine größere Anzahl von Items und betrachten mehrere Facetten Generalisierten Vertrauens. Insgesamt ergibt sich für die Objektiv-Skalen gegenüber der Standard-Skala damit eine differenziertere

Messung Generalisierten Vertrauens, die möglicherweise in einer überzufälligen Differenz der Skalensummenwerte resultiert.

Für die Signifikanzprüfungen innerhalb des *between-subject*-Designs zeichnet sich ein ähnliches Bild, wie bereits im Rahmen des *within-subject*-Designs. Innerhalb von Gruppe A ergeben sich für die Mehrzahl der untersuchten Paarungen sowohl bei der parametrischen als auch bei der nonparametrischen Signifikanzprüfung keine überzufälligen Unterschiede im Antwortverhalten auf den verschiedenen Vertrauensskalen. Für die verbleibenden zwei Fälle kann nur für PR_1 vs. PR_2 das bei der parametrischen Prüfung gefundene signifikante Ergebnis in der nonparametrischen Prüfung repliziert werden. Unter der Bedingung, dass die Messintention der Objektiv-Skala aufgedeckt wurde, stellen die Objektiv-Skalen und die Standard-Skala somit tendenziell äquivalente Messmethoden dar.

In Gruppe B ergeben sich hauptsächlich überzufällige Unterschiede zwischen den durchschnittlichen Summenwerten der Objektiv-Skalen und denen der Standard-Skala, die sich wechselseitig durch die parametrischen und nonparametrischen Signifikanzprüfungen bestätigen. Lediglich für die Kontraste PT_3 vs. PT_4 , $PT_{3(KS)}$ vs. PT_4 und $PR_{1(KS)}$ vs. PT_4 ergeben sich weder beim parametrischen noch beim nonparametrischen Test signifikante Unterschiede zwischen den Messmethoden. Wenn sich für die Items der Objektiv-Skalen eine geringe Augenscheinvalidität hinsichtlich der Zieldimension Generalisiertes Vertrauen realisieren lässt, stellen sich die Probanden auf den Objektiv-Skalen somit tendenziell weniger vertrauensvoll dar als auf der Standard-Skala. In jedem Fall sind die Objektiv-Skalen und die Standard-Skala aber wenigstens äquivalente Messmethoden.

Sowohl innerhalb des *within-subject*-Designs als auch innerhalb des *between-subject*-Designs ergeben sich für die Probandengruppe, die die Messintention der Objektiv-Skalen erkannt hat, mehrheitlich überzufällige Unterschiede im Antwortverhalten auf den verschiedenen Vertrauensskalen. Für die Probandengruppe, die die Messintention der Objektiv-Skalen erkannt hat, ergeben sich dagegen mehrheitlich keine überzufälligen Unterschiede zwischen den durchschnittlichen Skalensummenwerten. Hieraus lässt sich schließen, dass es sich bei den Objektiv-Skalen und der Standard-Skala um mindestens gleichwertige Methoden zur Erfassung Generalisierten Vertrauens handelt. Eine erfolgreiche Verschleierung der Messintention der Objektiv-Skalen sorgt allerdings dafür, dass die Probanden sich auf diesen Skalen weniger vertrauensvoll darstellen als auf der Standard-Skala. Ob dies einer Verbesserung der Messgüte entspricht bzw. ob die Probanden auf den Objektiv-Skalen aufrichtiger geantwortet haben als auf der Standard-Skala oder umgekehrt, kann an dieser Stelle jedoch noch nicht zweifelsfrei bestimmt werden.

7.2.6.9 Effekt des Treatments (Identifikation der Messintention und Referenzgruppe)

Strenggenommen wird sowohl durch die Objektiv-Skalen als durch die Standard-Skala nur dann tatsächlich Generalisiertes Vertrauen abgebildet, wenn die Probanden generalisierte Andere bzw. ihre *outgroup* als Referenzgruppe ihres Vertrauens angeben. Werden dagegen spezifische Andere bzw. die jeweilige *ingroup* seitens der Probanden als Bezugspunkt des Vertrauens genannt, so wird im Grunde Partikulares Vertrauen gemessen. Diese Problematik, auf welchen

Personenkreis sich das mittels Vertrauensskalen ermittelte Vertrauen bezieht, wird innerhalb der Literatur als Radiusproblem⁸⁷ der Vertrauensmessung bezeichnet.

Im folgenden Abschnitt soll diesem Umstand Rechnung getragen werden, indem in die Betrachtung der Differenz zwischen dem Antwortverhalten auf den Objektiv-Skalen und dem Antwortverhalten auf der Standard-Skala getrennt nach der Identifikation der Zieldimension der Objektiv-Skalen nur solche Fälle Eingang finden, die sowohl bezüglich ihrer Antworten auf den Objektiv-Skalen als auch bezüglich ihrer Antworten auf der Standard-Skala generalisierte Andere bzw. die *outgroup* als Vertrauensreferenz angegeben haben. Auf eine Betrachtung der Fälle, die spezifische Andere bzw. die *ingroup* als Vertrauensreferenz genannt haben, wurde verzichtet, da die Erforschung der Messung Partikularen Vertrauens nicht Gegenstand dieser Untersuchung sein soll.

Für die zu tätigen Analysen gelten darüber hinaus die gleichen Vorüberlegungen wie schon im vorangegangenen Abschnitt, in dem nur nach der Identifikation der Messintention der Objektiv-Skalen differenziert wurde. Ebenso werden innerhalb des *within-subject*- und *between-subject*-Designs die gleichen Messwertpaarungen einer Signifikanzprüfung unterzogen wie im vorigen Abschnitt. Außerdem wird gleichermaßen angenommen, dass in der Probandengruppe, welche die Messintention der Objektiv-Skalen nicht identifizieren konnte (Gruppe B), tendenziell mehr überzufällige Unterschiede zwischen den Skalensummenwerten der Objektiv-Skalen und der Standard-Skala auftreten, als in der Probandengruppe, die die Messintention der Objektiv-Skalen identifizieren konnte (Gruppe A). In Gruppe A sollten die Skalensummenwerte dagegen eher einen Trend zur Äquivalenz aufweisen.

7.2.6.9.1 *Within-Subject-Design*

Im Rahmen des *within-subject*-Designs werden zur Überprüfung der Kontraste ausschließlich nonparametrische Verfahren herangezogen. Dieser Schritt ist damit zu begründen, dass durch die obengenannten Auswahlkriterien, die in die Signifikanzprüfung eingehende Anzahl von Fällen stark reduziert wird ($N < 20$).

Als Prüfverfahren kommt der Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben und es werden getrennt nach Identifikation der Messintention (Gruppe A) und Nicht-Identifikation der Messintention (Gruppe B) jeweils die folgenden Hypothesen untersucht:

H_0 : Die Mediane der Objektiv-Skala bzw. -Kurzskala (PR_1 bzw. PT_2) zur Erfassung generalisierten und der Standard-Skala (PT_1 bzw. PR_2) zur Erfassung Generalisierten Vertrauens fallen in der Grundgesamtheit gleich aus bzw. die Rangsummen der beiden Messvarianten sind in der Population gleich. Unterschiede zwischen der Objektiv-Skala bzw. -Kurzskala und der Standard-Skala sind rein zufällig.

$$W^+ = W^- \text{ bzw. } \tilde{x}_{PR} = \tilde{x}_{PT1} \text{ bzw. } \tilde{x}_{PR2} = \tilde{x}_{PT2}$$

H_1 : Der Median der Objektiv-Skala bzw. -Kurzskala (PR_1) fällt innerhalb der Grundgesamtheit niedriger aus als der Median der Standard-Messung (PT_1). Zwischen den beiden Messvarianten treten positive Ränge mit größerer Häufigkeit auf als negative Ränge.

$$W^+ > W^- \text{ bzw. } \tilde{x}_{PR1} < \tilde{x}_{PT}$$

bzw.

⁸⁷ Vgl. Abschnitt „4.3.2 Radius generalisierten Vertrauens“

7. Hauptstudie

Der Median der Standard-Messung (PR_2) fällt innerhalb der Grundgesamtheit höher aus als der Median der Objektiv-Skala bzw. -Kurzskala (PT_2). Zwischen den beiden Messvarianten treten negative Ränge mit größerer Häufigkeit auf als positive Ränge.

$$W^+ < W^- \text{ bzw. } \tilde{x}_{PR_2} > \tilde{x}_{PT_2}$$

**Tab. 7.77: Ergebnisse der nonparametrischen Signifikanzprüfung
(Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben)**

Within-subject-Design [Messintention der Objektiv-Skalen identifiziert und Referenzgruppe entspricht <i>outgroup</i>]	
PR₁ vs. PT₁	$z=-2,449$ (einseitig); $p<.05$; $\varphi=.866$; $1-\beta=.690$
PR_{1(KS)} vs. PT₁	$z=-1,732$ (zweiseitig); n.s.; $\varphi=.612$; $1-\beta=.306$
PR₂ vs. PT₂	$z=-2,121$ (einseitig); $p<.05$; $\varphi=.588$; $1-\beta=.697$
PR₂ vs. PT_{2(KS)}	$z=-1,342$ (zweiseitig); n.s.; $\varphi=.372$; $1-\beta=.403$

z : empirischer z -Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p>.05$ (zweiseitig))

* $p<.05$ (einseitig), ** $p<.01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f;

Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 278

(Quelle: eigene Berechnungen)

In Gruppe A ergeben sich jeweils für die Gegenüberstellung von Objektiv-Skala und Standard-Skala (PR_1 vs. PT_1 und PR_2 vs. PT_2) signifikante Unterschiede im Antwortverhalten, während die Differenz zwischen Objektiv-Kurzskala und Standardskala jeweils nicht überzufällig ausfällt ($PR_{1(KS)}$ vs. PT_1 und PR_2 vs. $PT_{2(KS)}$). Für alle betrachteten Kontraste liegt die post-hoc-Teststärke allerdings unterhalb des notwendigen Niveaus von $1-\beta=.80$. Das bedeutet: Die signifikanten Ergebnisse können nicht statistisch abgesichert werden, während sich für die nicht-signifikanten Ergebnisse eine β -Fehler-Wahrscheinlichkeit oberhalb des Grenzwertes von $\beta=.20$ ergibt. Die letztgenannten Resultate sind somit nicht eindeutig interpretierbar bzw. die Datengrundlage zu gering, als dass eine Entscheidung zu Gunsten der H_1 gefällt werden konnte. (Bortz (2005), S. 122; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Folgt man den Werten der Effektstärken, so handelt es sich bei den Unterschieden zwischen den Objektiv-Skalen und der Standard-Skala um moderate bis starke Effekte ($\omega \geq .30$). Das besagt: Es existieren praktisch bedeutsame Unterschiede zwischen den verschiedenen Vertrauensskalen. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 270, S. 633; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Für die aus *Versuchsgruppe 1* stammenden Skalenpaarungen ergeben sich jeweils mehr Fälle für die das per Standard-Skala ermittelte Generalisierte Vertrauen höher ausfällt als das mittels der Objektiv-Skalen ermittelte Generalisierte Vertrauen. Die Kontrastierung von Objektiv-Kurzskala und Standard-Skala weist dabei allerdings eine erhebliche Anzahl von Fällen auf, für die sich kein Unterschied zwischen den beiden Vertrauensskalen ergibt. Die Rangsumme für positive Ränge fällt allerdings in beiden Fällen höher aus als die Rangsumme für negative Ränge. Die Richtung des potentiellen Unterschiedes entspricht somit dem Postulat der zugehörigen Alternativhypothese.

Für die aus *Versuchsgruppe 2* stammenden Skalenpaarungen ergeben sich ebenfalls jeweils mehr Fälle, für die das mittels der Objektiv-Skala ermittelte Generalisierte Vertrauen niedriger ausfällt als das per Standard-Skala ermittelte Generalisierte Vertrauen. Allerdings existiert sowohl für den Vergleich von Objektiv-Skala und Standard-Skala als auch für den Vergleich von

7. Hauptstudie

Objektiv-Kurzskala und Standard-Skala eine nicht unerhebliche Anzahl von Fällen, für die die Werte auf beiden Vertrauensskalen gleich groß ausfallen. Insgesamt fällt die Rangsumme für negative Ränge aber jeweils höher aus als die Rangsumme für positive Ränge, was insofern dem von der zugeordneten Alternativhypothese postulierten Verhältnis entspricht.

Ausgehend von der z-verteilten Prüfgröße ergeben sich sowohl aus der Differenz der Rangsummen von PR_1 vs. PT_1 als auch aus der Differenz der Rangsummen von PR_2 vs. PT_2 hochsignifikante Ergebnisse. Es wird dementsprechend jeweils die H_1 beibehalten und die H_0 abgelehnt. Die Probanden geben hier auf der Objektiv-Skala somit einen überzufällig niedrigeren Vertrauenslevel an als auf der Standard-Skala. Die übrigen Kontraste entsprechen zwar hinsichtlich der Richtung der Differenz dem Postulat der Alternativhypothese, allerdings resultiert kein überzufälliger Unterschied. Es werden demnach jeweils die H_1 abgelehnt und die H_0 beibehalten. Ausgehend von der zweiseitigen Signifikanzprüfung weicht damit das Messergebnis der Objektiv-Skalen weder nach unten noch nach oben vom Messergebnis der Standard-Skala ab und umgekehrt. Die Objektiv-Skalen und die Standard-Skala scheinen somit in Gruppe A mindestens äquivalente Messverfahren darzustellen.

Tab. 7.78: Ergebnisse der nonparametrischen Signifikanzprüfung (Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben)

Within-subject-Design [Messintention der Objektiv-Skalen <i>nicht</i> identifiziert und Referenzgruppe entspricht <i>outgroup</i>]	
PR_1 vs. PT_1	$z = -2,333$ (einseitig); $p < .05$; $\varphi = .602$; $1-\beta = .617$
$PR_{1(KS)}$ vs. PT_1	$z = -1,890$ (zweiseitig); n.s.; $\varphi = .488$; $1-\beta = .225$
PR_2 vs. PT_2	$z = -1,134$ (zweiseitig); n.s.; $\varphi = .275$; $1-\beta = .180$
PR_2 vs. $PT_{2(KS)}$	$z = -.378$ (zweiseitig); n.s.; $\varphi = .092$; $1-\beta = .064$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

1- β : Teststärke

vgl. Bühner/Ziegler (2009), S. 278

(Quelle: eigene Berechnungen)

In Gruppe B ergibt sich ausschließlich für die Skalenpaarung PR_1 vs. PT_1 ein überzufälliger Unterschied zwischen der Objektiv-Skala und der Standard-Skala. Die verbleibenden Skalenpaarungen weisen jeweils keine signifikanten Kontraste auf. Unabhängig davon kann aufgrund der Werte der post-hoc-Teststärke unterhalb des notwendigen Niveaus von $1-\beta = .80$ keines der Ergebnisse statistisch abgesichert werden. Vielmehr ergibt sich für die nicht-signifikanten Resultate eine erhöhte β -Fehler-Wahrscheinlichkeit ($\beta > .20$), sodass diese Ergebnisse nicht eindeutig interpretierbar sind. Die Datenbasis des Signifikanztests ist offensichtlich zu gering, um eine Entscheidung zu Gunsten der H_1 zu treffen. (Bortz (2005), S. 122; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Die Effektstärken zeigen Unterschiede zwischen den Objektiv-Skalen und der Standard-Skala an, die als kein Effekt bis, im Falle von PR_1 vs. PT_1 , zu einem starken Effekt reichen ($0 > \omega > .50$). Für die meisten der Kontrastierungen kann daher eine praktisch bedeutsame Differenz zwischen durchschnittlichen Skalensummenwerten angenommen werden. (Bortz (2005), S. 120; Bühner/Ziegler (2009), S. 177ff, S. 270, S. 633; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Die Skalenpaarungen aus *Versuchsgruppe 1* weisen jeweils eine höhere Rangsumme für positive Ränge als für negative Ränge auf, sodass der potentielle Kontrast im Antwortverhalten auf den verschiedenen Vertrauensskalen dem Postulat der zugehörigen Alternativhypothese entspricht. Innerhalb von *Versuchsgruppe 1* ist somit eine Tendenz dazu zu beobachten, dass mehr Fälle existieren, für die das per Objektiv-Skalen ermittelte Generalisierte Vertrauen niedriger ausfällt als das per Standard-Skala erhobene Generalisierte Vertrauen. Allerdings existiert auch eine Reihe von Fällen, für die das Antwortverhalten auf den Objektiv-Skalen und der Standard-Skala identisch ausfällt.

In *Versuchsgruppe 2* ergeben sich zumindest für PR_2 vs. PT_2 mehr Fälle, für die das mittels der Objektiv-Skala ermittelte Generalisierte Vertrauen niedriger ausfällt als das per Standard-Skala ermittelte Generalisierte Vertrauen. Für den Kontrast von Objektiv-Kurzskala und Standard-Skala fällt die Relation dagegen umgekehrt aus. Allerdings ergibt sich für beide Skalenpaarungen eine große Anzahl von Fällen, bei denen sich das Antwortverhalten auf den verschiedenen Vertrauensskalen nicht unterscheidet. Es scheint somit allenfalls der erstgenannte Kontrast der von der zugehörigen Alternativhypothese formulierten Richtung zu entsprechen.

Die z-verteilte Prüfgröße zeigt ausschließlich für den Unterschied zwischen PR_1 und PT_1 ein signifikantes Ergebnis an. Es wird daher die H_1 beibehalten und die H_0 verworfen. Die übrigen Kontraste wiesen bei einer zweiseitigen Signifikanzprüfung jeweils keine Überzufälligkeit auf. Es scheinen hier demnach keine wesentlichen Abweichungen zwischen dem Antwortverhalten auf den verschiedenen Vertrauensskalen zu existieren. Innerhalb von Gruppe B stellen die Objektiv-Skalen und die Standard-Skala somit mehrheitlich gleichwertige Messmethoden dar. Allerdings können aufgrund der niedrigen Teststärken falsch-nicht-signifikante Ergebnisse nicht ausgeschlossen werden. Ein gewichtiger Faktor sind dabei zusätzlich die geringen Fallzahlen, die in die einzelnen Signifikanzprüfungen eingehen.

7.2.6.9.2 Between-subject-Design

Im Rahmen des *between-subject*-Design werden sowohl parametrische als auch nonparametrische Tests durchgeführt. Allerdings gilt hier für ausgewählte Kontraste, dass die in die Analyse eingehende Fallzahl entweder deutlich asymmetrisch auf die verschiedenen Vertrauensskalen verteilt ist oder insgesamt vermindert ist. Dieser Umstand lässt sich insbesondere für die Skalenvergleiche innerhalb der Probandengruppe beobachten, welche die Messintention der Objektiv-Skalen identifizieren konnte (Gruppe A). Die Ergebnisse aus dieser Probandengruppe sind entsprechend kritisch zu bewerten.

7.2.6.9.2.1 Parametrische Signifikanzprüfung

Zunächst sollen die Ergebnisse der parametrischen Signifikanzprüfung betrachtet werden. Hierbei wurde der T-Test für unabhängige Stichproben eingesetzt.

Zu prüfen sind die folgenden Hypothesen, die sowohl für die Probandengruppe, welche die Messintention der Objektiv-Skalen erkannt hat (Gruppe A), als auch für die Probandengruppe, welche die Messintention der Objektiv-Skalen nicht erkannt hat (Gruppe B), getestet werden sollen:

H_0 : Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR_1 , PT_3) zur Erfassung Generalisierten Vertrauens unterscheiden sich jeweils nicht vom Mittelwert der

7. Hauptstudie

Standard-Skala (PR₂, PT₄) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwert ergibt null.

$$\mu_{PT3} = \mu_{PT4} \text{ oder } \mu_{PT3} - \mu_{PT4} = 0$$

$$\mu_{PR1} = \mu_{PR2} \text{ oder } \mu_{PR1} - \mu_{PR2} = 0$$

$$\mu_{PR1} = \mu_{PT4} \text{ oder } \mu_{PR1} - \mu_{PT4} = 0$$

$$\mu_{PR2} = \mu_{PT3} \text{ oder } \mu_{PR2} - \mu_{PT3} = 0$$

H₁: Die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR₁, PT₃) zur Erfassung Generalisierten Vertrauens fallen jeweils niedriger aus als der Mittelwert der Standard-Skala (PR₂, PT₄) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwerte ist kleiner als null.

$$\mu_{PT3} < \mu_{PT4} \text{ oder } \mu_{PT3} - \mu_{PT4} < 0$$

$$\mu_{PR1} < \mu_{PR2} \text{ oder } \mu_{PR1} - \mu_{PR2} < 0$$

$$\mu_{PR1} < \mu_{PT4} \text{ oder } \mu_{PR1} - \mu_{PT4} < 0$$

bzw.

Der Mittelwert der Standard-Skala (PR₂, PT₄) zur Erfassung Generalisierten Vertrauens fällt jeweils höher aus als die Mittelwerte der Objektiv-Skala bzw. -Kurzskala (PR₁, PT₃) zur Erfassung Generalisierten Vertrauens. Die Differenz der Populationsmittelwerte ist größer als null.

$$\mu_{PR2} > \mu_{PT3} \text{ oder } \mu_{PR2} - \mu_{PT3} > 0$$

Im Rahmen der Signifikanzprüfung ist zu beachten, dass für Gruppe A eine höhere Eintrittswahrscheinlichkeit für die H₀ angenommen wird, während für Gruppe B ein Trend zur H₁ erwartet wird.

Tab. 7.79: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Between-subject-Design [Messintention der Objektiv-Skalen identifiziert]	
PT ₃ vs. PT ₄	T(34)= ,000 (zweiseitig); n.s.; g=0; 1-β=,050
PR ₁ vs. PR ₂	T(22,714)= -2,730 (einseitig); p<,01; g=,711; 1-β=,596
PR ₁ vs. PT ₄	T(25,830)= -2,257 (einseitig); p<,05; g=,608; 1-β=,475
PR ₂ vs. PT ₃	T(36)= ,038 (zweiseitig); n.s.; g=,020; 1-β=,050

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<,05 (einseitig), **p<,01 (einseitig)

n.s.: nicht signifikant (p>,05 (zweiseitig))

g: Effektstärke mit d = ,20 → *kleiner Effekt*, d = ,50 → *mittlerer Effekt*, d = ,80 → *großer Effekt* (Eid/Gollwitzer/Schmitt (2015, S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

Tab. 7.80: Parametrische Verfahren (T-Test für unabhängige Stichproben)**Between-subject-Design (Kurzskala) [Messintention der Objektiv-Skalen identifiziert]**

PT_{3(KS)} vs. PT₄	T(34)= ,000 (zweiseitig); n.s.; $g=0$; $1-\beta=,050$
PR_{1(KS)} vs. PR₂	T(45)= -,687 (zweiseitig); n.s.; $g=,255$; $1-\beta=,103$
PR_{1(KS)} vs. PT₄	T(39)= -,729 (zweiseitig); n.s.; $g=,167$; $1-\beta=,072$
PR₂ vs. PT_{3(KS)}	T(36)= ,041 (zweiseitig); n.s.; $g=,021$; $1-\beta=,050$

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

* $p<,05$ (einseitig), ** $p<,01$ (einseitig)n.s.: nicht signifikant ($p>,05$ (zweiseitig)) g : Effektstärke mit $d = .20 \rightarrow$ *kleiner Effekt*, $d = .50 \rightarrow$ *mittlerer Effekt*, $d = .80 \rightarrow$ *großer Effekt* (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f) $1-\beta$: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

Stellt man die Mittelwerte der Objektiv-Skalen und der Standard-Skala innerhalb der Skalenpaarungen gegenüber, lässt sich beobachten, dass diese für den Skalenvergleich von PT₃ vs. PT₄ und PT_{3(KS)} vs. PT₄ gleich groß ausfallen, während für den Skalenvergleich von PR₂ vs. PT₃ und PR₂ vs. PT_{3(KS)} im Durchschnitt auf den Objektiv-Skalen leicht höhere Vertrauenswerte erreicht werden als auf der Standard-Skala. Die übrigen Kontraste entsprechen hinsichtlich der Relation der Skalensummenwerte dem Postulat der Alternativhypothese.

Innerhalb von Gruppe A ergaben sich ausschließlich für die Gegenüberstellung von PR₁ vs. PR₂ und PR₁ vs. PT₄ überzufällige Unterschiede zwischen dem Antwortverhalten auf der Objektiv-Skala und dem Antwortverhalten auf der Standard-Skala. Alle übrigen Kontrastierungen weisen keine signifikanten Ergebnisse auf. Für alle Signifikanzprüfungen gilt allerdings, dass die jeweiligen Ergebnisse nicht statistisch abgesichert werden können. Für die nicht-signifikanten Ergebnisse resultiert jeweils eine extrem hohe β -Fehler-Wahrscheinlichkeit ($\beta>,20$), sodass an dieser Stelle davon ausgegangen werden muss, dass die Datenbasis der Signifikanzprüfungen zu klein war, um eine Entscheidung zu Gunsten der H₁ treffen zu können. Die geringen Teststärken sind jedoch nicht zuletzt auch der geringen Fallzahlen der jeweiligen Objektiv-Stichproben geschuldet. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Signifikante Varianzunterschiede zwischen den in die Analyse eingehenden Stichproben lassen sich ausschließlich für die Kontraste PR₁ vs. PR₂ und PR₁ vs. PT₄ beobachten. Diese sind dabei so gelagert, dass konservative Testentscheidungen begünstigt werden. Aufgrund des jeweils überzufälligen Unterschiedes scheint dies allerdings keinen Einfluss auf die Testentscheidung gehabt zu haben. (Bortz (2005), S. 141)

Die Effektstärken nehmen je nach Skalenpaarungen Werte an, die keinem Effekt bis hin zu einem mittleren Effekt entsprechen ($.20 < g < .80$). Die Unterschiede zwischen den Objektiv-Skalen und der Standard-Skala scheinen demnach nur für ausgesuchte Skalenpaarungen praktische Relevanz zu besitzen. Das 95%-Konfidenzintervall der Differenz überstreicht in allen Fällen, mit Ausnahme von PR₁ vs. PR₂, den Wert null, sodass mehrheitlich mit einer großen Wahrscheinlichkeit davon ausgegangen werden muss, dass tatsächlich Nulldifferenzen zwischen den verschiedenen Vertrauensskalen vorliegen. In Wahrheit unterscheidet sich das Antwortverhalten somit eher nicht. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

7. Hauptstudie

Für die Kontraste von PR_1 vs. PR_2 und PR_1 vs. PT_4 wird vorläufig die H_1 beibehalten und die H_0 abgelehnt, wenn auch aufgrund des Wertebereichs des 95%-Konfidenzintervall der Differenz für PR_1 vs. PT_4 und der post-hoc-Teststärke unterhalb des notwendigen Schwellenwertes von $1-\beta=.80$ zu vermuten ist, dass es sich um ein falsch-signifikantes Ergebnis handelt.

Für die übrigen Kontraste wird jeweils die H_0 beibehalten und die H_1 abgelehnt. Es kann somit davon ausgegangen werden, dass die Objektiv-Skalen und die Standard-Skala innerhalb von Gruppe A vornehmlich äquivalente Messverfahren darstellen.

Tab. 7.81: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Between-subject-Design [Messintention der Objektiv-Skalen nicht identifiziert]	
PT_3 vs. PT_4	$T(56,580) = -1,476$ (zweiseitig); n.s.; $g=.377$; $1-\beta=.294$
PR_1 vs. PR_2	$T(51,269) = -1,491$ (zweiseitig); n.s.; $g=.378$; $1-\beta=.261$
PR_1 vs. PT_4	$T(49,293) = -1,382$ (zweiseitig); n.s.; $g=.347$; $1-\beta=.223$
PR_2 vs. PT_3	$T(59) = -1,559$ (zweiseitig); n.s.; $g=.402$; $1-\beta=.335$

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

* $p<.05$ (einseitig), ** $p<.01$ (einseitig)

n.s.: nicht signifikant ($p>.05$ (zweiseitig))

g : Effektstärke mit $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

$1-\beta$: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

Tab. 7.82: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Between-subject-Design (Kurzskala) [Messintention der Objektiv-Skalen nicht identifiziert]	
$PT_{3(KS)}$ vs. PT_4	$T(56,580) = -1,095$ (zweiseitig); n.s.; $g=.280$; $1-\beta=.184$
$PR_{1(KS)}$ vs. PR_2	$T(51,549) = -.867$ (zweiseitig); n.s.; $g=.219$; $1-\beta=.119$
$PR_{1(KS)}$ vs. PT_4	$T(49,003) = -.836$ (zweiseitig); n.s.; $g=.209$; $1-\beta=.111$
PR_2 vs. $PT_{3(KS)}$	$T(59) = -1,135$ (zweiseitig); n.s.; $g=.293$; $1-\beta=.201$

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

* $p<.05$ (einseitig), ** $p<.01$ (einseitig)

n.s.: nicht signifikant ($p>.05$ (zweiseitig))

g : Effektstärke mit $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

$1-\beta$: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

(Quelle: eigene Berechnungen)

In Gruppe B folgt die Relation der durchschnittlichen Skalensummenwerte für alle Skalenpaarungen dem Postulat der Alternativhypothese. Die Probanden stellen sich demnach auf den Objektiv-Skalen jeweils weniger vertrauensvoll dar als auf der Standard-Skala.

Unabhängig davon stellen sich innerhalb von Gruppe B für keinen der Kontraste signifikante Unterschiede im Antwortverhalten auf den verschiedenen Vertrauensskalen ein. Gleichzeitig liegt allerdings die β -Fehler-Wahrscheinlichkeit oberhalb des Grenzwertes von $\beta=.20$ und die post-hoc-Teststärke unterhalb des notwendigen Niveaus von $1-\beta=.80$. Es kann somit nicht ausgeschlossen werden, dass es sich um falsch-nicht-signifikante Ergebnisse handelt, vielmehr ist die Wahrscheinlichkeit, eine richtige H_1 anzunehmen, stark reduziert. Möglicherweise ist die Datenbasis der Signifikanzprüfungen zu klein, als dass eine Entscheidung zu Gunsten der H_1 gefällt werden könnte. (Bortz (2005), S. 121ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 188ff; Cohen (1988), S. 4; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

Mit Ausnahme von PR_2 vs. PT_3 und PR_2 vs. $PT_{3(KS)}$ ergeben sich jeweils signifikante Varianzunterschiede zwischen den in die jeweiligen Signifikanzprüfungen eingehenden Stichproben. Dies deutet daraufhin, dass auf den Objektiv-Skalen möglicherweise anders geantwortet wurde als auf der Standard-Skala. Zudem begünstigt die Relation der Varianzen konservative Testentscheidungen, da die Varianz in der kleineren Stichprobe (hier: Objektiv-Messung) jeweils kleiner ausfiel als in der größeren Stichprobe (hier: Standard-Messung). Vor diesem Hintergrund sind die nicht-signifikanten Unterschiede kritisch zu hinterfragen bzw. es erhöht sich für die Paarungen die Gefahr von falsch-nicht-signifikanten Ergebnissen. (Bortz (2005), S. 141)

Die Effektstärken zeigen abhängig von der Skalenpaarung kleine oder mittlere Unterschiede an. Es handelt sich somit um praktisch bedeutsame Differenzen zwischen den Vertrauensskalen. Allerdings enthält das 95%-Konfidenzintervall der Differenz für jede der Paarungen auch den Wert null, sodass auch eine Nulldifferenz im Bereich des Möglichen liegt. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 819ff; Bühner/Ziegler (2009), S. 179ff, S. 258, S. 265; Cohen (1988), S. 9f, S. 20ff, S. 66ff; Eid/Gollwitzer/Schmitt (2015), S. 337f)

Einstweilen werden für alle Kontraste die H_0 beibehalten und die H_1 verworfen. Hierbei gilt es allerdings zu bedenken, dass eine Tendenz zu konservativen Testentscheidungen, d.h. eine Bevorzugung der H_0 , besteht, die Fehler-Wahrscheinlichkeit vergleichsweise hoch ausfällt und die Effektstärken relativ stark von null abweichen. Die Gefahr von falsch-nicht-signifikanten Ergebnissen ist somit nicht ausgeräumt. Allerdings stellen die Objektiv-Skalen und die Standard-Skala in Gruppe B zumindest gleichwertige Methoden zur Erfassung Generalisierten Vertrauens dar.

7.2.6.9.2.2 Nonparametrische Signifikanzprüfung

Zum Zweck der Validierung der Ergebnisse der parametrischen Signifikanzprüfungen werden nachfolgend die gleichen Kontraste mit Hilfe nonparametrischer Verfahren untersucht. Hierzu wird der Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben eingesetzt.

Die innerhalb der Probandengruppe, welche die Messintention der Objektiv-Skalen erkannt hat (Gruppe A), und der Probandengruppe, welche die entsprechende Messintention nicht erkannt hat (Gruppe B), zu prüfenden Kontraste lassen sich in den folgenden Hypothesen zusammenfassen:

H_0 : Die Mediane der Objektiv-Skala bzw. -Kurzskala zur Erfassung Generalisierten Vertrauens (PR_1 , PT_3) unterscheiden sich nicht vom Median der Standard-Skala zur Erfassung Generalisierten Vertrauens (PR_2 , PT_4). In der Grundgesamtheit liegen gleich viele Rangplatzüberschreitungen wie -unterschreitungen vor.

$$\tilde{x}_{PT} = \tilde{x}_{PT4}$$

$$\tilde{x}_{PR1} = \tilde{x}_{PR}$$

$$\tilde{x}_{PR} = \tilde{x}_{PT}$$

$$\tilde{x}_{PT} = \tilde{x}_{PR}$$

H_1 : Die Mediane der Objektiv-Skala bzw. -Kurzskala zur Erfassung Generalisierten Vertrauens (PR_1 , PT_3) sind kleiner als der Median der Standard-Skala zur Erfassung

7. Hauptstudie

Generalisierten Vertrauens (PR_2 , PT_4). In der Grundgesamtheit liegen mehr Rangplatzunterschreitungen als Rangplatzüberschreitungen vor.

$$\tilde{x}_{PT3} < \tilde{x}_{PT4}$$

$$\tilde{x}_{PR1} < \tilde{x}_{PR2}$$

$$\tilde{x}_{PR1} < \tilde{x}_{PT4}$$

$$\tilde{x}_{PT3} < \tilde{x}_{PR2}$$

Auch für die Signifikanzprüfung mit Hilfe der nonparametrischen Verfahren wird angenommen, dass die H_0 in Gruppe A eine höhere Eintrittswahrscheinlichkeit gegenüber der H_1 besitzt, während für Gruppe B das Gegenteil, d.h. eine höhere Eintrittswahrscheinlichkeit der H_1 gegenüber der H_0 , angenommen wird.

**Tab. 7.83: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design [Messintention der Objektiv-Skalen identifiziert und Referenzgruppe entspricht <i>outgroup</i>]	
PT₃ vs. PT₄	$z = -.054$ (zweiseitig); n.s.; $\varphi = .009$; $1-\beta = .050$
PR₁ vs. PR₂	$z = -1.884$ (einseitig); $p < .05$; $\varphi = .287$; $1-\beta = .182$
PR₁ vs. PT₄	$z = -1.668$ (einseitig); $p < .05$; $\varphi = .260$; $1-\beta = .163$
PR₂ vs. PT₃	$z = -.129$ (zweiseitig); n.s.; $\varphi = .021$; $1-\beta = .050$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

**Tab. 7.84: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design (Kurzskala) [Messintention der Objektiv-Skalen identifiziert und Referenzgruppe entspricht <i>outgroup</i>]	
PT_{3(KS)} vs. PT₄	$z = -.108$ (zweiseitig); n.s.; $\varphi = .018$; $1-\beta = .050$
PR_{1(KS)} vs. PR₂	$z = -.439$ (zweiseitig); n.s.; $\varphi = .067$; $1-\beta = .053$
PR_{1(KS)} vs. PT₄	$z = -.453$ (zweiseitig); n.s.; $\varphi = .071$; $1-\beta = .054$
PR₂ vs. PT_{3(KS)}	$z = -.079$ (zweiseitig); n.s.; $\varphi = .013$; $1-\beta = .050$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

Innerhalb von Gruppe A ergeben sich ausschließlich für die Skalenpaarungen PR_1 vs. PR_2 und PR_1 vs. PT_4 überzufällige Unterschiede im Antwortverhalten. Unabhängig davon kann jedoch keines der Ergebnisse statistisch abgesichert werden. Die post-hoc-Teststärke nimmt jeweils Werte unterhalb von $1-\beta = .80$ an. Gleichzeitig fällt die β -Fehler-Wahrscheinlichkeit sehr hoch aus. Falsch-nicht-signifikante Ergebnisse können demnach nicht ausgeschlossen werden. Die niedrigen Teststärken sind dabei allerdings wohl der kleinen Fallanzahl geschuldet, die auf Seiten der Objektiv-Skalen in die Signifikanzprüfungen eingeht. (Bortz (2005), S. 122ff; Bortz/Döring (2016), S. 669ff; Bühner/Ziegler (2009), S. 191; Cohen (1988), S. 3ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 222ff; Sedlmeier/Renkewitz (2013), S. 383ff, S. 729f)

7. Hauptstudie

Die Effektstärken fallen vergleichsweise niedrig aus und zeigen maximal kleine Differenzen zwischen den durchschnittlichen Skalensummenwerten der Objektiv-Skalen und der Standard-Skala an ($\omega < .30$). Die Differenzen zwischen den Messmethoden besitzen demnach kaum praktische Relevanz. (Bortz (2005), S. 120; Bortz/Döring (2016), S. 668f; Bühner/Ziegler (2009), S. 177ff, S. 269f, S. 633; Cohen (1988), S. 9f, S. 223ff; Cohen (1992), S. 156; Eid/Gollwitzer/Schmitt (2015), S. 228ff)

Ausgehend vom Verhältnis der mittleren Ränge entsprechen in Gruppe A alle Skalenpaarungen mit Ausnahme von PT_3 vs. PT_4 und PR_2 vs. PT_3 der Vorhersage der Alternativhypothese. In der Mehrheit der Fälle erreichen die Probanden, deren Generalisiertes Vertrauen anhand der Objektiv-Skalen erfasst wurde, niedrigere Ränge als die Probanden, deren Generalisiertes Vertrauen mit Hilfe der Standard-Skala gemessen wurde.

Ausgehend von den z-verteilten Prüfgrößen wird für die Kontraste PR_1 vs. PR_2 und PR_1 vs. PT_4 die H_1 beibehalten und die H_0 abgelehnt. Allerdings ist dieses Resultat aufgrund der niedrigen Teststärken und der in die Analyse eingehenden Fallzahlen kritisch zu betrachten. Ein falsch-signifikantes Ergebnis kann an dieser Stelle nicht ausgeschlossen werden. (Bühner/Ziegler (2009), S. 287)

Des Weiteren werden für die übrigen Kontraste jeweils die H_0 beibehalten und die H_1 verworfen. Innerhalb von Gruppe A sind die verschiedenen Messmethoden somit als eher äquivalent anzusehen.

**Tab. 7.85: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design [Messintention der Objektiv-Skalen <u>nicht</u> identifiziert und Referenzgruppe entspricht <i>outgroup</i>]	
PT_3 vs. PT_4	$z = -1,447$ (zweiseitig); n.s.; $\varphi = ,188$; $1-\beta = ,106$
PR_1 vs. PR_2	$z = -1,471$ (zweiseitig); n.s.; $\varphi = ,200$; $1-\beta = ,104$
PR_1 vs. PT_4	$z = -1,320$ (zweiseitig); n.s.; $\varphi = ,183$; $1-\beta = ,094$
PR_2 vs. PT_3	$z = -1,575$ (zweiseitig); n.s.; $\varphi = ,202$; $1-\beta = ,117$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

**Tab. 7.86: Ergebnisse der nonparametrischen Signifikanzprüfung
(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)**

Between-subject-Design (Kurzskala) [Messintention der Objektiv-Skalen <u>nicht</u> identifiziert und Referenzgruppe entspricht <i>outgroup</i>]	
$PT_{3(KS)}$ vs. PT_4	$z = -1,106$ (zweiseitig); n.s.; $\varphi = ,144$; $1-\beta = ,083$
$PR_{1(KS)}$ vs. PR_2	$z = -,963$ (zweiseitig); n.s.; $\varphi = ,131$; $1-\beta = ,073$
$PR_{1(KS)}$ vs. PT_4	$z = -,889$ (zweiseitig); n.s.; $\varphi = ,123$; $1-\beta = ,070$
PR_2 vs. $PT_{3(KS)}$	$z = -1,178$ (zweiseitig); n.s.; $\varphi = ,151$; $1-\beta = ,087$

z: empirischer z-Wert

(einseitig) bzw. (zweiseitig): Testrichtung

n.s.: nicht signifikant ($p > .05$ (zweiseitig))

* $p < .05$ (einseitig), ** $p < .01$ (einseitig)

φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

$1-\beta$: Teststärke

vgl. Bühner/Ziegler (2009), S. 286

(Quelle: eigene Berechnungen)

Innerhalb von Gruppe B ergeben sich keinerlei signifikante Unterschiede zwischen den unterschiedlichen Messskalen. Allerdings gehen die nicht-signifikanten Unterschiede mit einer sehr hohen β -Fehler-Wahrscheinlichkeit einher. Es ist somit anzunehmen, dass es sich um durchweg falsch-nicht-signifikante Ergebnisse handelt bzw. möglicherweise die Datenbasis der Signifikanzprüfungen zu klein ausfiel, als dass zu Gunsten der H_1 entschieden werden konnte. Aus den post-hoc-Teststärken lässt sich außerdem ablesen, dass die Wahrscheinlichkeit, eine korrekte H_1 anzunehmen, stark reduziert ist.

Die Effektstärken zeigen zwar praktisch relevante, allerdings allenfalls kleine Differenzen zwischen den Messmethoden an ($\omega < .30$).

Betrachtet man die mittleren Ränge, so ergeben sich in der Mehrzahl der Paarungen für die Probanden, deren Generalisiertes Vertrauen mit Hilfe der Objektiv-Skala erfasst wurde, im Durchschnitt niedrigere Ränge als für die Probanden, deren Generalisiertes Vertrauen anhand der Standard-Skala erfasst wurde. Die Richtung des Unterschieds zwischen beiden Messverfahren entspricht somit dem von der Alternativhypothese vorausgesagten Trend. Eine Ausnahme hiervon bildet die Kontrastierung von PR_1 und PT_4 . Hier verläuft die Relation umgekehrt, was darauf hindeutet, dass die Probanden sich auf der Objektiv-Skala vertrauensvoller darstellen als auf der Standard-Skala. (Bühner/Ziegler (2009), S. 287)

Vorläufig wird für alle betrachteten Kontraste innerhalb von Gruppe B die H_0 beibehalten und H_1 abgelehnt. Aufgrund der geringen Teststärken und der hohen β -Fehler-Wahrscheinlichkeit ist Gefahr falsch-nicht-signifikanter Ergebnisse allerdings vergleichsweise hoch. Die Objektiv-Skalen und die Standard-Skala stellen jedoch zumindest gleichwertige Messverfahren dar.

7.2.6.9.3 Zusammenfassung

Im vorangegangenen Abschnitt wurden die Objektiv-Skalen und die Standard-Skala unter den Bedingungen gegeneinander getestet, dass die Zieldimension der Objektiv-Skalen nicht identifiziert wurde und bei der Bearbeitung der Skalen jeweils die *outgroup* als Vertrauensreferenz angegeben wurde. Weiterhin wurde angenommen, dass innerhalb der Probandengruppe, die die Messintention der Objektiv-Skalen nicht identifizieren konnte, eine größere Differenz zwischen Messskalen eintritt, als in der Probandengruppe, welche die entsprechende Messintention identifizieren konnte.

Zunächst kann festgestellt werden, dass die durchschnittliche Differenz zwischen den Summenscores der Objektiv-Skalen und der Standard-Skala unabhängig von der Identifikation der Messintention und unabhängig vom zugrunde gelegten experimentellen Design nicht größer als .75 Skalenpunkte ausfällt. Der Unterschied zwischen den Messskalen ist somit relativ gering. Es scheint demnach zunächst kaum einen Unterschied zu machen, welche Skala man zur Erfassung des Vertrauenslevels verwendet. Die Analyse zeigt jedoch, dass zumindest im Rahmen des *between-subject*-Design die Unterschiede innerhalb von Gruppe B (keine Identifikation der Messintention) größer ausfallen als in Gruppe A (Identifikation der Messintention), auch wenn die Differenzen nicht überzufällig sind.

Im Rahmen des *within-subject*-Design ergeben sich zwischen den Signifikanzprüfungen von Gruppe A und B kaum Unterschiede. Dies scheint hauptsächlich der geringen Fallzahlen geschuldet zu sein, die in die jeweiligen Signifikanzprüfungen eingehen. Insbesondere für Gruppe A fällt die Stichprobengröße sehr klein aus, sodass schon geringe Unterschiede einen großen Effekt haben. Entsprechend ergeben sich innerhalb Gruppe A bei zwei von vier Kontrastierungen signifikante Ergebnisse, während in Gruppe B nur in einem von vier Fällen ein

signifikantes Ergebnis resultiert. Ausgehend hiervon scheint die Identifikation der Messintention der Objektiv-Skalen eher keinen Einfluss auf das Antwortverhalten der Probanden zu haben. Die Resultate der Signifikanzprüfungen deuten dagegen eher daraufhin, dass es sich bei den verschiedenen Messskalen um äquivalente Messmethoden handelt.

Im Rahmen des *between-subject*-Design konnten sowohl in Gruppe A als auch in Gruppe B die signifikanten Ergebnisse der parametrischen Signifikanzprüfung durch die signifikanten Ergebnisse der nonparametrischen Signifikanzprüfung bestätigt werden. Allerdings gehen auch hier vornehmlich in Gruppe A auf Seiten der Objektiv-Skalen teilweise nur wenige Fälle in die Analyse ein, sodass extreme Antworten hier einen besonders großen Effekt haben können.

Entsprechend ergeben sich in Gruppe A für die Kontraste PR_1 vs. PR_2 und PR_1 vs. PT_4 signifikante Ergebnisse, während die übrigen Kontraste keine überzufälligen Unterschiede aufweisen. Dabei weisen allerdings alle Kontraste niedrige post-hoc-Teststärken auf, die im Falle der Skalenpaarungen mit nicht-signifikanten Differenzen zwischen des Summenscores gegen null tendieren. Gleiches gilt für die Effektstärken, die nur marginale Unterschiede anzeigen. Insgesamt lässt dies vermuten, dass die signifikanten Ergebnisse eher durch Zufall zustande gekommen sind bzw. aus der Kennwertverteilung von PR_1 resultieren. So enthält PR_1 unter den oben genannten Bedingungen lediglich 9 Fälle, die sich zwar in der Mehrzahl auf die Mittelkategorie verteilen, wobei darüber hinaus aber nur Fälle auf der linken Seite der Skala zu finden sind. Hieraus resultiert ein im Vergleich sehr niedriger Mittelwert, der für den analyseübergreifend größten Abstand zwischen Objektiv- und Standard-Skala sorgt. Daraus wird geschlossen, dass unter der Bedingung, die Messintention der Objektiv-Skala erkannt zu haben, in Wirklichkeit Äquivalenz zwischen den unterschiedlichen Messskalen besteht. Diese These wird zusätzlich dadurch unterstützt, dass sich innerhalb von Gruppe A mehrheitlich nicht-signifikante Ergebnisse einstellen. Die Objektiv-Skalen und die Standard-Skala sind dann als gleichwertige Methoden zur Erfassung Generalisierten Vertrauens zu betrachten.

In Gruppe B ergeben sich zwar weder im Zuge der parametrischen noch im Zuge der nonparametrischen Signifikanzprüfungen überzufällige Unterschiede zwischen den verschiedenen Messskalen, allerdings fallen die Effektstärken jeweils deutlich höher aus als in Gruppe A und das Varianzverhältnis der in die Analyse eingehenden Stichproben lässt eine Neigung zu konservativen Testentscheidungen vermuten. Auch ist die β -Fehler-Wahrscheinlichkeit deutlich erhöht. Zudem entspricht das Verhältnis der durchschnittlichen Skalensummenwerte für jede der Skalenpaarungen der von der H_1 postulierten Relation. Somit stellen sich die Probanden unter der Bedingung, dass die Messintention der Objektiv-Skala nicht erkannt wurde, auf den Objektiv-Skalen weniger vertrauensvoll dar als auf der Standard-Skala. Für die Skalenpaarungen in Gruppe A konnte dies nicht beobachtet werden. Insgesamt lässt sich somit trotz allem eine Tendenz zu einer Differenz im Antwortverhalten zwischen den Objektiv-Skalen und der Standard-Skala feststellen.

Es lässt sich nun schließen, dass sowohl unter den experimentellen Bedingungen des *within-subject*-Design als auch des *between-subject*-Design, sofern die Probanden generalisierte Andere bzw. die *outgroup* als Vertrauensbezugspunkt angeben, unabhängig von der Identifikation der Messintention die Objektiv-Skalen und die Standard-Skala mindestens äquivalente Messverfahren für die Abbildung Generalisierten Vertrauens darstellen. Ferner lässt sich ein belastbarer Trend erkennen, wonach die Nicht-Identifikation der Zieldimension der Objektiv-Skalen bewirkt, dass sich die Probanden auf den Objektiv-Skalen weniger vertrauensvoll darstellen als auf der Standard-Skala. Ob diese Tendenz einem aufrichtigeren

Antwortverhalten bzw. einem Antwortverhalten frei vom Einfluss sozialer Erwünschtheit entspricht, kann an dieser Stelle noch nicht festgelegt werden.

7.2.6.10 Fazit der Signifikanzprüfungen

Im vorangegangenen Abschnitt wurde mit Hilfe parametrischer und nonparametrischer Verfahren untersucht, ob sich zwischen der Messung Generalisierten Vertrauens nach der Methode Objektiver Tests und der Standardmethode zur Erfassung Generalisierten Vertrauens ein signifikanter Unterschied ergibt. Zudem wurde geprüft, ob dieser potentielle Unterschied zwischen den Messmethoden durch Effekte der experimentellen Situation (*Pretest*-, *Positions*- oder *Carry-over*-Effekte) beeinflusst bzw. herbeigeführt sein könnte oder ob der Kontrast durch die Art und Weise des Messens bewirkt wurde. Die simultane Verwendung verteilungsbasierter und verteilungsfreier Tests dient dabei für beide Analysen zur wechselseitigen Validierung der Ergebnisse.

7.2.6.10.1 Effekte der experimentellen Situation

Im Rahmen der Prüfung auf potentielle Messverzerrungen durch Effekte der experimentellen Situation wurde geprüft, ob sich die einzelnen Objektiv-Messungen signifikant voneinander unterscheiden bzw. ob sich die einzelnen Standard-Messungen signifikant voneinander unterscheiden. Messungen, denen eine Befragung vorangegangen war, unabhängig davon ob es sich um die Objektiv- oder die Standard-Skala handelte, mussten dabei als mit einem Bias belastet angenommen werden. Dagegen konnten Messungen, denen keine Befragung vorangegangen war, als *one-shot-case-study* interpretiert werden und daher Biasfreiheit angenommen werden. Als potentiell biasbelastet wurde daher auf Seiten der Objektiv-Skala die Messung in PT₂ bzw. PT_{2(KS)} eingestuft, während auf Seiten der Standard-Skala die Messung in PT₁ als potentiell mit einem Bias belastet gewertet wurde. Als frei von einem Bias können dagegen PR₁ bzw. PR_{1(KS)} und PT₃ bzw. PT_{3(KS)} für die Objektiv-Skala und PR₂ und PT₄ für die Standard-Skala angesehen werden.

Damit eine Beeinflussung des Unterschiedes zwischen Objektiv- und Standard-Skala zur Erfassung Generalisierten Vertrauens durch etwaige *Pretest*-, *Positions*- und *Carry-over*-Effekte ausgeschlossen werden konnte, durfte sich das Antwortverhalten bei Messungen, denen bereits eine Befragung vorausgegangen war, nicht vom Antwortverhalten bei Messungen, denen keine Befragung vorangegangen war, unterscheiden. Ein signifikanter Unterschied im Antwortverhalten würde dagegen anzeigen, dass die jeweiligen Vorher-Messungen einen Einfluss auf die Probanden ausgeübt haben.

Betrachtet man das Ergebnis der parametrischen Analyse (hier: einfaktorielle Varianzanalyse) so ergibt sich weder für die Objektiv-Skala, die Objektiv-Kurzskala noch die Standard-Skala ein signifikanter Unterschied zwischen den Versuchsgruppen. Die Effektstärke nimmt Werte von $.095 \leq f \leq .105$ an, was einem kleinen Effekt entspricht. Der Kontrast zwischen den Gruppen ist entsprechend schwach ausgeprägt. Ausgehend von der post-hoc-Teststärke, die Werte von $.168 \leq 1-\beta \leq .192$ annimmt, kann die Übereinstimmung des Antwortverhaltens über alle Versuchsgruppen hinweg allerdings zunächst nicht statistisch abgesichert werden. Das hierfür erforderliche Teststärkeniveau von $1-\beta=.80$ wird stark unterschritten bzw. die Wahrscheinlichkeit zugunsten der Alternativhypothese (H_1) zu entscheiden, wenn diese in der

7. Hauptstudie

Grundgesamtheit Gültigkeit besitzt, ist zu gering. Voraussichtlich ist die in die Analyse eingehende Fallzahl zu gering, um eine eindeutige Interpretation des Testresultates zu erlauben.

Tab. 7.87: Parametrische Verfahren

Pretest-, Positions- und Carry-over-Effekt	
PR₁ vs. PT₂ vs. PT₃	F(2;149)=,838; n.s.; f=,105; 1-β=,192
PR_{1(KS)} vs. PT_{2(KS)} vs. PT_{3(KS)}	F(2;149)=,838; n.s.; f=,105; 1-β=,192
PT₁ vs. PR₂ vs. PT₄	F(2;149)=,708; n.s.; f=,095; 1-β=,168

F(df₁;df₂): empirischer F-Wert inklusive Zähler- und Nennerfreiheitsgrade
n.s.: nicht signifikant (p>.05 (zweiseitig))
f: Effektstärke $f = .10 \rightarrow$ kleiner Effekt; $f = .25 \rightarrow$ moderater Effekt; $f = .40 \rightarrow$ großer Effekt
1-β: Teststärke
Vgl. Eid/Gollwitzer/Schmitt (2015), S. 407
(Quelle: eigene Berechnungen)

Um trotz der geringen Fallzahl eine belastbare Aussage hinsichtlich der Beeinflussung der Messungen durch Effekte der experimentellen Situation machen zu können, wurde zusätzlich anhand verteilungsfreier Verfahren geprüft, ob sich das Antwortverhalten auf den Objektiv- bzw. Standard-Skalen über die Gruppen hinweg unterscheidet.

Auch die nonparametrische Analyse (hier: Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben) kommt zu dem Ergebnis, dass sich weder für die Objektiv-Skala, die Objektiv-Kurzskala noch die Standard-Skala ein signifikanter Unterschied zwischen den Versuchsgruppen ergibt. Die Effektstärke bewegt sich mit $.026 \leq \varphi \leq .114$ auf einem ähnlichen Niveau wie bereits bei der parametrischen Analyse. Der Kontrast zwischen den Versuchsgruppen kann somit als marginal bezeichnet werden. Die post-hoc-Teststärke rangiert mit $.083 \leq 1-\beta \leq .307$ weit unterhalb des notwendigen Niveaus von $1-\beta=.80$. Für sich genommen kann somit auch anhand der nonparametrischen Verfahren das Resultat der Signifikanzprüfung nicht statistisch abgesichert werden.

Tab. 7.88: Nonparametrische Verfahren

(Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test für unabhängige Stichproben)

Pretest-, Positions- und Carry-over-Effekt	
PR₁ vs. PT₂	z=-,799 (einseitig); n.s.; $\varphi=.080$; 1-β=,198
PR_{1(KS)} vs. PT_{2(KS)}	z=-,143 (einseitig); n.s.; $\varphi=.114$; 1-β=,307
PT₂ vs. PT₃	z=-,533 (einseitig); n.s.; $\varphi=.053$; 1-β=,133
PT_{2(KS)} vs. PT_{3(KS)}	z=-,532 (einseitig); n.s.; $\varphi=.053$; 1-β=,133
PT₁ vs. PR₂	z=-,264 (einseitig); n.s.; $\varphi=.026$; 1-β=,083
PT₁ vs. PT₄	z=-,509 (einseitig); n.s.; $\varphi=.050$; 1-β=,126

z: empirischer z-Wert
(einseitig) bzw. (zweiseitig): Testrichtung
n.s.: nicht signifikant (p>.05 (zweiseitig))
* $p<.05$ (einseitig), ** $p<.01$ (einseitig)
 φ : Effektstärke mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)
1-β: Teststärke
vgl. Bühner/Ziegler (2009), S. 286
(Quelle: eigene Berechnungen)

Die Überprüfung durch die nonparametrischen Verfahren bestätigt das parametrische Testergebnis, was die endgültige Beibehaltung der Nullhypothese (H_0) zur Folge hat. (Bortz/Lienert/Boehnke (2008), S. 85f) Es wird demnach davon ausgegangen, dass hinsichtlich des Antwortverhaltens auf den unterschiedlichen Vertrauensskalen kein signifikanter Unterschied zwischen den Versuchsgruppen besteht. Daraus lässt sich schließen, dass die Resultate der Messungen in PT₂, PT_{2(KS)} und PT₁ nicht von den Resultaten ihrer jeweiligen als frei von einem Bias angenommenen Pendanten abweichen. Außerdem kann nun davon ausgegangen

werden, dass innerhalb des Surveyexperimentes keine Beeinflussung der verschiedenen Messungen durch *Pretest*-, *Positions*- oder *Carry-over*-Effekte stattgefunden hat. Dies stellt gleichzeitig einen Hinweis darauf dar, dass die im Rahmen der Objektiv-Messung realisierte geringe Augenscheinvalidität hinsichtlich der Zieldimension „Generalisiertes Vertrauen“ wirksam war.

7.2.6.10.2 Auswirkungen des Treatment

Zur Bestimmung des Effektes des Treatments wurden die Objektiv-Skala und die Standard-Skala sowie die Objektiv-Kurzskala und die Standard-Skala unter zwei experimentellen Bedingungen (*within-subject*- und *between-subject*-Design) kontrastiert. Um zu prüfen, ob sich Unterschiede im Antwortverhalten zwischen den Objektiv-Skalen und der Standard-Skala ergeben, wurden dabei Signifikanztests auf drei verschiedenen Differenzierungsebenen durchgeführt. Grundlegende Annahme hierbei war, dass sich die Probanden auf den Objektiv-Skalen jeweils weniger vertrauensvoll präsentieren als auf der Standard-Skala. Zusatzannahme der höheren Differenzierungsebenen war außerdem, dass sich unter der Bedingung der Nicht-Identifikation der Messintention der Objektiv-Skalen größere Unterschiede bzw. mehr überzufällige Differenzen zwischen den durchschnittlichen Skalensummenwerten der Objektiv-Skalen und der Standard-Skala ergeben sollten, als wenn die entsprechende Messintention identifiziert werden konnte. In letzterem Fall sollten die Objektiv-Skalen und die Standard-Skala gleichwertige Messverfahren darstellen. Das heißt: Es sollte eine Tendenz zu äquivalenten Skalensummenwerten bestehen.

Auf der ersten Differenzierungsebene gingen alle Fälle in die Analyse ein, die je nach Experimentalgruppe einen gültigen Wert auf den Objektiv-Skalen und der Standard-Skala (*Experimentalgruppen 1 und 2*) hatten oder auf einem von beiden Skalentypen einen gültigen Wert (*Experimentalgruppen 3 und 4*) hatten. Die Differenzierung erfolgte hierbei entlang der Designgrenzen des Surveyexperimentes, sodass die Versuchsgruppen entweder als *within-subject*-Design oder als *between-subject*-Design angeordnet waren. Hierbei ergaben sich in der Mehrzahl signifikante Unterschiede im Antwortverhalten zwischen den Objektiv-Skalen und der Standard-Skala. Diese entsprachen dabei der Forderung, dass auf den Objektiv-Skalen tendenziell ein niedrigeres Vertrauenslevel angegeben wird als auf der Standard-Skala. Sowohl die signifikanten als auch die nicht-signifikanten Ergebnisse konnten wechselseitig durch parametrische und nonparametrische Signifikanzprüfungen repliziert werden. Im *within-subject*-Design und im *between-subject*-Design fielen dabei die Unterschiede zwischen der Objektiv-Skala und der Standard-Skala jeweils größer aus als zwischen der Objektiv-Kurzskala und der Standard-Skala. In Kombination mit den Resultaten der Bestimmung der Konstruktvalidität lässt sich daraus schließen, dass die Objektiv-Kurzskala eine größere inhaltliche Ähnlichkeit zur Standard-Skala aufweist als die Objektiv-Skala. Aus den wenigen nicht-signifikanten Ergebnissen lässt sich indes ablesen, dass es sich bei den verschiedenen Messverfahren mindestens um äquivalente Messmethoden handelt. Für die mehrheitlich signifikanten Ergebnisse können dagegen zwei konkurrierende Interpretationen herangezogen werden. Auf der einen Seite könnte die Differenz zwischen den Messskalen daher rühren, dass die Probanden auf mindestens einer Skala weniger aufrichtig antworten. Entsprechend der hier zugrunde gelegten Hypothesen, wird angenommen, dass für das Antwortverhalten auf der Standard-Skala ein größerer Einfluss reaktiver Effekte zu erwarten ist. Auf der anderen Seite könnte sich die beschriebene Differenz auch aus der inhaltlichen Verfasstheit der Messskalen

ableiten. So fragen die Objektiv-Skalen Vertrauen differenzierter und anhand einer größeren Anzahl von Items ab als die Standard-Skala. Dies könnte Einfluss auf die für die jeweilige Abfrage angenommenen *demand characteristics* nehmen. In der Folge wird womöglich für die Objektiv-Skalen eine andere bzw. leicht abweichende Zieldimension angenommen als für die Standard-Skala, sodass ein Unterschied im Antwortverhalten resultiert. Aufgrund der Tatsache, dass die Objektiv-Skalen und die Standard-Skala stark miteinander korrelieren, wird dieser zweite Erklärungsansatz allerdings verworfen.

Auf der zweiten Differenzierungsebene gingen ebenfalls alle Fälle, die auf den verschiedenen Vertrauensskalen einen gültigen Wert besaßen in die Analyse ein. Auch wurde wieder zwischen *within-subject*-Design und *between-subject*-Design differenziert. Zusätzlich wurden die Probanden innerhalb jedes der experimentellen Designs in zwei Probandengruppen aufgeteilt, wobei eine diejenigen Personen enthielt, welche die Messintention der Objektiv-Skalen identifizieren konnten (Gruppe A), während in der anderen Gruppe, diejenigen Personen zusammengefasst wurden, welche die Messintention der Objektiv-Skalen nicht identifizieren konnten (Gruppe B). Hierbei ergaben sich sowohl innerhalb des *within-subject*-Design als auch innerhalb des *between-subject*-Design für Gruppe A mehr nicht-signifikante Ergebnisse als signifikante Ergebnisse, während für Gruppe B mehr signifikante als nicht-signifikante Ergebnisse festgestellt werden konnten. Mit anderen Worten: Unter der Bedingung, dass die Messintention nicht erkannt wurde, ergaben sich tendenziell eher überzufällige Unterschiede im Antwortverhalten zwischen den Objektiv-Skalen und der Standard-Skala. In allen Fällen fiel dabei der Vertrauenslevel auf den Objektiv-Skalen niedriger aus als auf der Standard-Skala. Dagegen erwiesen sich die Objektiv-Skalen und die Standard-Skala unter der Bedingung, dass die Messintention der Objektiv-Skalen erkannt wurde, mehrheitlich als tendenziell gleichwertige Messverfahren. Dies lässt den Schluss zu, dass sich aus der erfolgreichen Verschleierung der Zieldimension der Objektiv-Skalen eine weniger vertrauensvolle Selbstdarstellung auf den Objektiv-Skalen im Vergleich zur Standard-Skala ergibt. Ob dies eine Verbesserung der Messgüte gleichkommt bzw. dies Folge einer Verringerung des Einflusses reaktiver Effekte, im Speziellen sozialerwünschten Antwortverhaltens, ist, kann an dieser Stelle jedoch nicht zweifelsfrei bestimmt werden. Es bleibt somit zunächst unklar, ob die Probanden auf den Objektiv-Skalen aufrichtiger geantwortet haben als auf der Standard-Skala oder umgekehrt.

Die dritte Differenzierungsebene entspricht im Wesentlichen der zweiten Differenzierungsebene, allerdings gehen in die Analyse nur noch Fälle ein, die sowohl für ihr Antwortverhalten auf den Objektiv-Skalen als auch auf der Standard-Skala generalisierte Andere bzw. die *outgroup* als Vertrauensreferenz angegeben haben. Entsprechend des Radiusproblems der Vertrauensforschung wird genau genommen nur dann Generalisiertes Vertrauen gemessen, wenn die letztgenannte Bedingung erfüllt ist. Werden dagegen spezifische Andere bzw. die *ingroup* als Vertrauensreferenz verwendet, wird bei genauerer Betrachtung Partikulares Vertrauen erfasst. Im Rahmen des *within-subject*-Designs gab auf dieser Differenzierungsebene keine nachhaltigen Unterschiede zwischen Gruppe A und Gruppe B. Für signifikante Unterschiede scheint insbesondere in Gruppe A vornehmlich die geringe Fallzahl verantwortlich zu sein, die in die jeweiligen Analysen eingeht. Wenige extreme Antworten haben hier bereits einen großen Effekt auf die Differenz zwischen den durchschnittlichen Skalensummenwerten der Objektiv-Skalen und der Standard-Skala. Außerdem ergeben sich insgesamt mehr nicht-signifikante als signifikante Ergebnisse. Für die im *within-subject*-Design angeordneten Skalenpaarungen wird deshalb angenommen, dass die Identifikation der Messintention eher keinen Einfluss auf das Antwortverhalten auf den Objektiv-Skalen hat bzw. es sich bei den

Objektiv-Skalen und der Standard-Skala tendenziell um mindestens gleichwertige Messverfahren zur Erfassung Generalisierten Vertrauens handelt. Im Rahmen des *between-subject*-Designs ergeben sich ausschließlich innerhalb von Gruppe A signifikante Unterschiede zwischen den Messskalen. Allerdings scheint auch hier wieder die Kombination aus niedrigen Fallzahlen und verhältnismäßig extremen Antworten für die überzufällige Differenz verantwortlich zu sein. Für Gruppe B ergeben sich keinerlei signifikante Unterschiede zwischen den durchschnittlichen Skalensummenwerten der Objektiv-Skalen und dem der Standard-Skala, was auch durch nonparametrische Signifikanzprüfungen repliziert werden konnte. Allerdings ergeben sich mehrheitlich Varianzunterschiede, was auf Abweichungen im Antwortverhalten zwischen den Skalentypen hindeutet. Unabhängig von der Signifikanz gilt dabei für alle Vergleiche, dass sich die Probanden auf der Objektiv-Skala weniger vertrauensvoll darstellen als auf der Standard-Skala. Die Verschleierung der Messintention der Objektiv-Skalen scheint demnach in der Tendenz den gleichen Effekt zu haben, wie er bereits für die zweite Differenzierungsebene berichtet wurde.

Insgesamt lassen sich für die hier durchgeführten Analysen nun zwei Interpretationsalternativen heranziehen. Aufgrund der Tatsache, dass die signifikanten Ergebnisse für den Vergleich der Objektiv-Skalen mit der Standard-Skala über die verschiedenen Differenzierungsebenen insbesondere in Gruppe B nicht stabil bleiben, besteht die Möglichkeit, dass auf den Differenzierungsebenen 2 und 3 das Radiusproblem eine Moderatorenrolle einnimmt. So kommen auf der ersten und zweiten Differenzierungsebene womöglich nur deshalb signifikante Unterschiede zwischen den Messskalen zustande, weil im Grunde zwei unterschiedliche Formen des Vertrauens abgefragt werden, die miteinander interferieren. Mit anderen Worten: Das Vertrauenslevel der Personen, die spezifische Andere als Referenzgruppe angegeben haben, unterscheidet sich möglicherweise fundamental vom Vertrauenslevel der Personen, die generalisierte Andere als Referenzgruppe angegeben haben, bzw. es existiert eine Differenz zwischen dem durchschnittlichen Partikularen und dem durchschnittlichen Generalisierten Vertrauen. Auf der dritten Differenzierungsebene wird dieser Effekt ausgeschaltet, woraufhin der Unterschied zwischen dem Antwortverhalten auf den Objektiv-Skalen und dem Antwortverhalten auf der Standard-Skala verschwindet.

Gegen diese These spricht zunächst, dass sich zwischen den Durchschnittswerten Partikularen Vertrauens und Generalisierten Vertrauens weder innerhalb noch zwischen den einzelnen Versuchsgruppen eine überzufällige Differenz ergibt⁸⁸. Partikulares und Generalisiertes Vertrauen üben somit wechselseitig keine Verstärkungs- oder Verminderungseffekte aufeinander aus. Außerdem zeigt sich, dass insbesondere für Gruppe B die Relation der durchschnittlichen Skalensummenwerte der Objektiv-Skalen und der Standard-Skala immer der gleichen Richtung folgt. So erreichen die Probanden auf den Objektiv-Skalen unter der Bedingung, dass deren Zieldimension erfolgreich maskiert wurde, grundsätzlich einen niedrigeren Vertrauenslevel als auf der Standard-Skala. Dieser Trend lässt sich für Gruppe A nicht beobachten. Zusätzlich ergeben sich auf der zweiten Differenzierungsebene mehrheitlich signifikante Ergebnisse in Gruppe B, während in Gruppe A mehrheitlich nicht-signifikante Ergebnisse auftreten. Und auf der ersten Differenzierungsebene lässt sich ein genereller Trend dazu entdecken, dass eine überzufällige Differenz zwischen den Objektiv-Skalen und der Standard-Skala existiert. Zusammengefasst deutet dies daraufhin, dass die Unterschiede zwischen den Messmethoden aus deren Merkmalen resultieren. Im Speziellen bedeutet das,

⁸⁸ Vgl. Anhang III: Partikulares vs. Generalisiertes Vertrauen

dass die geringe Augenscheinvalidität der Items der Objektiv-Skalen bzw. die erfolgreiche Verschleierung der Zieldimension der Objektiv-Skalen für die Skalensummenwertunterschiede verantwortlich zu sein scheint. Wird die Messintention nicht identifiziert, so ergibt sich auf den Objektiv-Skalen ein überzufällig niedrigerer Vertrauenslevel als auf der Standard-Skala. Ob dies einer Reduktion reaktiver Effekte und im Besonderen einer Verminderung des Einflusses sozialer Erwünschtheit entspricht bzw. ob dies bedeutet, dass das Antwortverhalten auf den Objektiv-Skalen aufrichtiger ausfällt, lässt sich an dieser Stelle jedoch noch nicht feststellen. In letzter Instanz sind die Objektiv-Skalen und die Standard-Skala allerdings mindestens äquivalente Herangehensweisen für die Messung Generalisierten Vertrauens.

7.2.7 Einfluss sozialer Erwünschtheit

Sozialerwünschte Antworten können sich in Situationen ergeben, die eine bezüglich der situationalen Rahmung besonders positive Selbstdarstellung erfordern, die bei unvorteilhafter Performanz negative Konsequenzen nach sich ziehen können oder innerhalb derer man ein positives Bild von sich selbst aufrechterhalten möchte. Soziale Erwünschtheit („*social desirability*“, SD) beschreibt somit ein Verhalten, dessen Ziel es ist, situationsspezifisch günstige bzw. erstrebenswerte Eigenschaften und Merkmale der eigenen Person hervorzuheben, während gleichzeitig negativ besetzte Eigenschaften und Merkmale verborgen werden sollen. Was als positiv oder negativ wahrgenommen wird, ist individuell, gruppenbezogen oder situational verschieden. Allerdings wird sich vornehmlich an allgemein anerkannten Normen und Werten und insbesondere an gesamtgesellschaftlich als wünschenswert geltendem Verhalten orientiert.

Ob die Antworten bezüglich interessierender Konstrukte durch soziale Erwünschtheit verzerrt wurden, kann anhand der Korrelationen dieser Konstrukte mit einer SD-Skala bestimmt werden. Dabei gilt: Je stärker der Zusammenhang zwischen SD-Skala und Konstrukt, desto stärker auch die Tendenz zu sozial erwünschtem Antworten in Bezug auf das Konstrukt. (Mummendey/Grau (2014), S. 178ff) Um den Einfluss sozialer Erwünschtheit auf die Angabe Generalisierten Vertrauens zu untersuchen, wird daher der Zusammenhang zwischen den verschiedenen Summenscores der BIDR-Kurzskala und den Summenscores der verschiedenen Vertrauensskalen betrachtet. Wie bereits beschrieben, differenziert die BIDR-Kurzskala mit Selbsttäuschung (*self-deceptive enhancement*, SDE) und Fremdtäuschung (*impression management*, IM) zwei voneinander zu unterscheidende Beweggründe für sozialerwünschtes Antwortverhalten, denen auch in der folgenden Analyse Rechnung getragen werden soll. Außerdem soll danach differenziert werden, ob die Probanden die Messintention der Objektiv-Skalen identifizieren konnten (Gruppe A) oder nicht (Gruppe B). In einem weiteren Differenzierungsschritt wird zusätzlich dem Radiusproblem Rechnung getragen, indem ausschließlich Fälle in die Analyse eingehen, die generalisierte Andere als Vertrauensreferenz angegeben haben. Eine Verbesserung der Messgüte bzw. der Validität der Messung infolge der Verwendung der Objektiv-Skalen kann dann unter den folgenden Bedingungen erwartet werden:

- (1) Die Objektiv-Skalen sind weniger stark mit der SD-Skala korreliert als die Standard-Skala. Das heißt: Die Objektiv-Skalen sind im Allgemeinen weniger anfällig für sozialerwünschtes Antwortverhalten.

- (2) Innerhalb von Gruppe B sind die Objektiv-Skalen weniger stark mit der SD-Skala korreliert als in Gruppe A. Das heißt: Die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen verringert deren Anfälligkeit für sozialerwünschtes Antwortverhalten.
- (3) Die Korrelation der Messwerte der Standard-Skala mit der SD-Skala fällt in Gruppe A und B ähnlich hoch aus. Das heißt: Die mit der Zieldimension übereinstimmende Augenscheinvalidität der Standard-Skala macht diese anfällig für sozialerwünschtes Antwortverhalten.
- (4) In Gruppe A sind die Objektiv-Skalen und die Standard-Skala ähnlich stark mit der SD-Skala korreliert. Das heißt: Die Transparenz der Zieldimension bei beiden Skalen bewirkt eine übereinstimmende Anfälligkeit der Skalen für sozialerwünschtes Antwortverhalten.
- (5) In Gruppe B sind die Objektiv-Skalen weniger stark mit der SD-Skala korreliert als die Standard-Skala. Das heißt: Die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen macht diese im Vergleich mit der Standard-Skala weniger anfällig für sozialerwünschtes Antwortverhalten.

Hierbei steht vornehmlich der Betrag der jeweiligen Korrelation mit der BIDR-Kurzskala und deren Subkomponenten im Fokus des Interesses.

Darüber hinaus soll auch betrachtet werden, inwiefern sozialerwünschtes Antwortverhalten einen Effekt auf die zwischen den Skalensummenwerten der Objektiv-Skalen und der Standard-Skala existierende Differenz hat. Hierzu sollen die folgenden Bedingungen geprüft werden:

- (1) Die Differenzen zwischen den Objektiv-Skalen und der Standard-Skala korreliert mit der SD-Skala. Das heißt: Sozialerwünschtes Antwortverhalten kann als Erklärungsvariable für den Unterschied zwischen den Objektiv-Skalen und der Standard-Skala herangezogen werden.
- (2) Die Differenzen zwischen den Objektiv-Skalen und der Standard-Skala korreliert sowohl in Gruppe A als auch in Gruppe B mit der SD-Skala. Das heißt: Die Differenz zwischen den Objektiv-Skalen und der Standard-Skala ergibt sich unabhängig von der Verschleierung der Messintention der Objektiv-Skalen durch sozialerwünschtes Antwortverhalten.
- (3) In Gruppe A korrelieren die Differenzen zwischen den Objektiv-Skalen und der Standard-Skala negativ mit der SD-Skala. Das heißt: Wenn die Messintention beider Skalentypen offenliegt, reduziert sich der Unterschied zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala infolge sozialerwünschten Antwortverhaltens.
- (4) In Gruppe B korrelieren die Differenzen zwischen den Objektiv-Skalen und der Standard-Skala positiv mit der SD-Skala. Das heißt: Wenn die Messintention der Objektiv-Skalen erfolgreich verschleiert werden konnte und nur die Messintention der Standard-Skala einsichtig ist, vergrößert sich der Unterschied zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala infolge sozialerwünschten Antwortverhaltens.
- (5) In Gruppe A fallen die Korrelationen der Einzelskalen mit der SD-Skala ähnlich stark aus, weisen allerdings einander entgegengesetzte Vorzeichen auf, während die SD-Skala negativ mit den Differenzen zwischen den Objektiv-Skalen und der Standard-Skala korreliert ist. Das heißt: Die Reduktion des Unterschiedes zwischen den Skalentypen ist darauf zurückzuführen, dass sowohl die Objektiv-Skalen als auch die Standard-Skala durch sozialerwünschtes Antwortverhalten beeinflusst sind. Mit anderen Worten: Die Angleichung der Messwerte ist darauf zurückzuführen, dass entweder die Objektiv-Skalen oder die Standard-Skala das wahre Vertrauen unterschätzen, während ihre Pendants das wahre Vertrauen überschätzen.
- (6) In Gruppe B korrelieren die Objektiv-Skalen weniger stark mit der SD-Skala als die Standard-Skala, während die SD-Skala positiv mit den Differenzen zwischen den Objektiv-Skalen und

der Standard-Skala korreliert ist. Das heißt: Die Vergrößerung des Unterschiedes zwischen den Skalentypen ist darauf zurückzuführen, dass vornehmlich die Standard-Skala durch sozialerwünschtes Antwortverhalten beeinflusst ist. Mit anderen Worten: Die Ungleichheit der Messwerte ist darauf zurückzuführen, dass die Standard-Skala das wahre Vertrauen überschätzt, während auf den Objektiv-Skalen aufrichtiger geantwortet wird.

7.2.7.1 Gruppe 1: Objektiv/Standard

Zunächst ist festzustellen, dass die Probanden in *Versuchsgruppe 1* eine leicht überdurchschnittliche Neigung zu sozialerwünschtem Antwortverhalten aufweisen ($\bar{x}_{BIDR} = 4,67$, $s = 0,922$; $\bar{x}_{SDE} = 4,71$, $s = 1,155$; $\bar{x}_{IM} = 4,45$, $s = 1,243$). Wobei die grafische Analyse einen starken Trend zu sozialerwünschtem Antwortverhalten erkennen lässt. Diese Tendenz ist sowohl für die Neigung zu *social desirability* im Allgemeinen als auch für deren Teilkomponenten Selbsttäuschung und Fremdtäuschung zu beobachten.

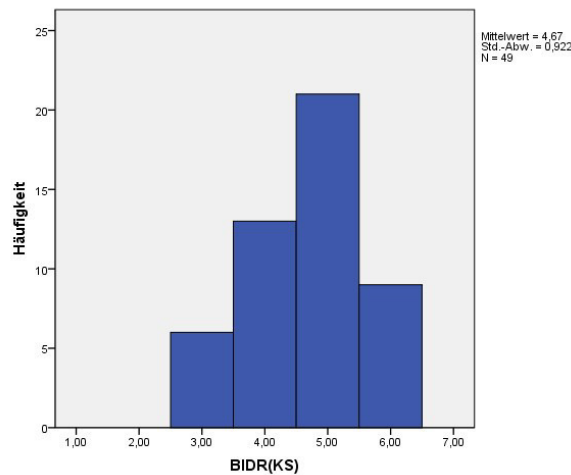


Abb. 7.25: Häufigkeitsverteilung Summenscore BIDR-Kurzskala („soziale Erwünschtheit“) (mit 1 = geringe Neigung zu SD bis 5 = starke Neigung zu SD) in Versuchsgruppe 1

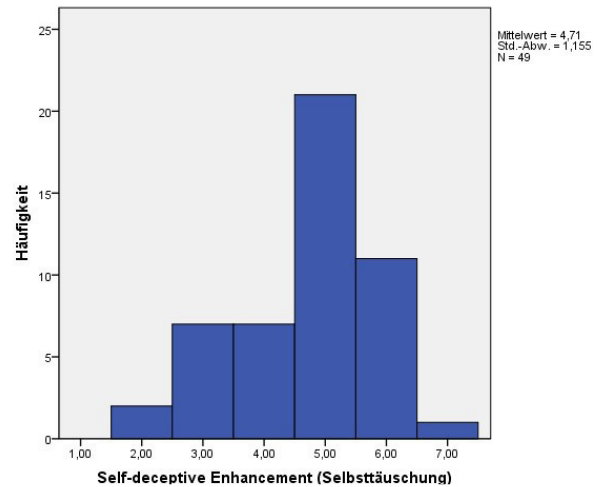


Abb. 7.26: Häufigkeitsverteilung Summenscore SDE-Komponente („Selbsttäuschung“) (mit 1 = geringe Neigung zu SDE bis 5 = starke Neigung zu SDE) in Versuchsgruppe 1

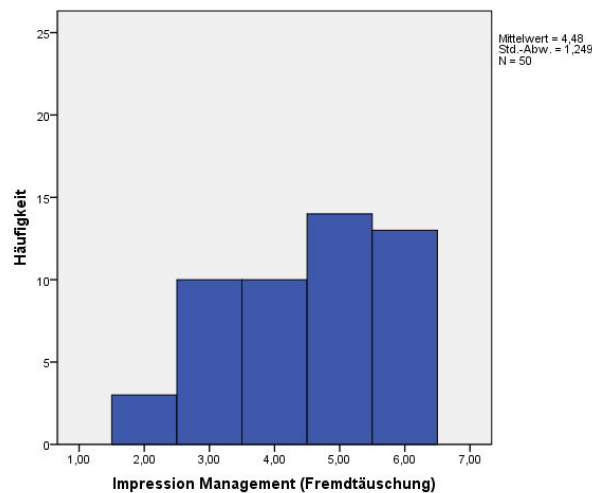


Abb. 7.27: Häufigkeitsverteilung Summenscore IM-Komponente („Fremdtäuschung“) (mit 1 = geringe Neigung zu IM bis 5 = starke Neigung zu IM) in Versuchsgruppe 1

Tab. 7.89: Bivariate Korrelationen: BIDR_(KS) und PR₁ bzw. PT₁ (N = 49)

	PR ₁	PR _{1(KS)}	PT ₁	BIDR _(KS)	SDE	IM
PR ₁	1	---	---	---	---	---
PR _{1(KS)}	,798**	1	---	---	---	---
PT ₁	,606**	,682**	1	---	---	---
BIDR _(KS)	,346*	,175	,276	1	---	---
SDE	,113	,056	,136	,615**	1	---
IM	,245	,078	,136	,695**	,019	1

PR₁: Objektiv-Skala

PR_{1(KS)}: Objektiv-Kurzskala

PT₁: Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Betrachtet man die Korrelationen der SD-Skala mit den verschiedenen Vertrauensskalen, fällt auf, dass sich lediglich schwache bis mittlere Zusammenhänge ergeben. Insgesamt ist daher anzunehmen, dass die Beeinflussung der Vertrauensmessung durch sozialerwünschtes Antwortverhalten eher weniger stark ausgeprägt ist.

Dabei sticht allerdings die Objektiv-Skala hervor. Hier ergibt sich ein signifikanter, und im Vergleich der stärkste, Zusammenhang mit der SD-Skala. Das Antwortverhalten auf der Objektiv-Skala scheint im Skalenvergleich demnach am stärksten durch soziale Erwünschtheit beeinflusst zu sein. Die Standard-Skala folgt auf dem zweiten Platz und die Objektiv-Kurzskala ist am wenigsten von sozialerwünschtem Antwortverhalten betroffen.

Insgesamt bewirkt die Neigung zu sozialerwünschtem Antwortverhalten allerdings auf jeder der Skalen einen Anstieg der Vertrauenswerte. Mit anderen Worten: Je größer die Neigung zu sozialerwünschtem Antwortverhalten, desto höher fällt das Generalisierte Vertrauen aus. Ausgehend davon, dass es sich bei einer stark ausgeprägten Vertrauensneigung um ein von der

7. Hauptstudie

Gesellschaft erstrebenswertes Gut handelt, ist daher anzunehmen, dass das wahre Generalisierte Vertrauen der Probanden geringer ausfällt, als die Angaben auf den Vertrauensskalen suggerieren.

Betrachtet man die Einzelkomponenten sozialer Erwünschtheit, so ist das Ergebnis weitestgehend deckungsgleich mit den Resultaten von NAEF/SCHUPP. So wirkt sich auch in *Versuchsgruppe 1* die IM-Komponente stärker auf das Antwortverhalten auf den Vertrauensskalen aus als die SDE-Komponente. (Naef/Schupp (2009), S. 10) Für die Standard-Skala sind die Anteile allerdings indifferent. Die Probanden scheinen daher zumindest bei der Beantwortung der Objektiv-Items stärker an einer Fremdtäuschung als an einer Selbsttäuschung interessiert gewesen zu sein. Sowohl der Einfluss von Selbsttäuschung und der Einfluss von Fremdtäuschung entsprechen dabei der anhand der Vertrauentheorie postulierten Richtung. So geht ein gesteigertes Generalisiertes Vertrauen gleichermaßen mit einer in optimistischer Weise verzerrten Weltwahrnehmung wie mit einer Tendenz zu einer möglichst guten bzw. vorteilhaften Außendarstellung der eigenen Person einher.

Augenscheinlich ermöglicht die Objektiv-Skala im Vergleich zur Standard-Skala in *Versuchsgruppe 1* keine Verbesserung der Validität der Vertrauensmessung. Tatsächlich ist sie sogar stärker von sozialerwünschten Antworten betroffen als die Standard-Skala. Anders sieht das für die aus der Objektiv-Skala abgeleitete Kurzskala aus. Die Objektiv-Kurzskala ist insgesamt deutlich weniger von sozialer Erwünschtheit betroffen. Begutachtet man die Korrelationen mit der SDE- und der IM-Komponente, lässt sich schließen, dass weder Selbsttäuschung noch Fremdtäuschung eine Rolle für das Antwortverhalten auf der Objektiv-Kurzskala spielen. Im Vergleich zur Standard-Skala stellt die Vertrauensmessung per Objektiv-Kurzskala somit aller Voraussicht nach eine Validitätssteigerung für die Vertrauensmessung dar.

Es kann somit geschlossen werden, dass zwar die Objektiv-Kurzskala, nicht aber die Objektiv-Skala weniger stark mit der SD-Skala korreliert ist als die Standard-Skala. Mit anderen Worten: Von den Objektiv-Skalen ist nur die Objektiv-Kurzskala weniger anfällig für sozialerwünschtes Antwortverhalten als die Standard-Skala.

Wie im Rahmen der Mittelwertvergleiche gezeigt werden konnte, unterscheiden sich die Summenscores von Objektiv- bzw. Objektiv-Kurzskala und Standard-Skala innerhalb von *Versuchsgruppe 1* überzufällig. Dabei zeigte sich, dass die Probanden auf den Objektiv-Skalen im Durchschnitt ein geringeres Generalisiertes Vertrauen angaben als auf der Standard-Skala. Nachfolgend soll nun untersucht werden, ob diese Differenz im Zusammenhang mit dem Einfluss sozialer Erwünschtheit steht.

Tab. 7.90: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 49)

	Diff (PT ₁ - PR ₁)	Diff (PT ₁ - PR _{1(KS)})	BIDR _(KS)	SDE	IM
Diff (PT ₁ - PR ₁)	1	---	---	---	---
Diff (PT ₁ - PR _{1(KS)})	,783**	1	---	---	---
BIDR _(KS)	,045	,169	1	---	---
SDE	,072	,118	,615**	1	---
IM	-,042	,093	,695**	,019	1

Diff (PT₁ - PR₁): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PT₁ - PR_{1(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Aus den Korrelationswerten lässt sich ablesen, dass zwischen sozialer Erwünschtheit und der Differenz der Summenscores der verschiedenen Vertrauensskalen ein geringer bis kein Zusammenhang besteht. Es ist daher anzunehmen, dass der Einfluss sozialer Erwünschtheit nicht bzw. nur zu einem sehr geringen Teil für die überzufällige Differenz zwischen den Objektiv-Skalen und der Standard-Skala verantwortlich ist. Dies gilt insbesondere für den Unterschied von Objektiv- und Standard-Skala.

Es lässt sich daher schließen, dass der Unterschied zwischen Objektiv-Skala und Standard-Skala nicht durch den Einfluss sozialerwünschten Antwortverhaltens zustande kommt, während der Unterschied zwischen Objektiv-Kurzskala und Standard-Skala zumindest zu einem geringen Teil auf sozialerwünschtes Antwortverhalten zurückzuführen ist.

Die Differenz zwischen den Objektiv-Skalen und der Standard-Skala ist positiv mit der BIDR-Kurzskala korreliert.

Es lässt sich daher schließen, dass sich der Unterschied zwischen den Messwerten der verschiedenen Vertrauensskalen infolge sozialerwünschten Antwortverhaltens steigert.

Die Abstandszunahme zwischen den Vertrauensskalen geht dabei auf Seiten der Standard-Skala mit einer positiven Korrelation mit der BIDR-Kurzskala einher. Dabei fällt nur die Korrelation der Objektiv-Kurzskala mit der BIDR-Kurzskala niedriger als die Korrelation der Standard-Skala mit der BIDR-Kurzskala aus.

Es kann somit geschlossen werden, dass im Falle der Objektiv-Kurzskala die Abstandsvergrößerung zwischen den Vertrauensskalen dadurch zustande kommt, dass die Probanden ihr Vertrauen auf der Standard-Skala überschätzen, während sie auf der Objektiv-Kurzskala tendenziell aufrichtiger antworten. Für die Abstandsvergrößerung zwischen den Messwerten von Objektiv- und Standard-Skala ist dagegen vornehmlich das Antwortverhalten auf der Objektiv-Skala verantwortlich.

7.2.7.1.1 Differenzierung nach Identifikation der Messintention

Differenziert man innerhalb von *Versuchsgruppe 1* den Einfluss sozialer Erwünschtheit auf die verschiedenen Vertrauensskalen anhand der Personen, die die Messintention der Objektiv-Skalen erkannt haben bzw. nicht erkannt haben, ergeben sich teilweise erhebliche Veränderungen, was die Tendenz zu sozialerwünschtem Antworten auf diesen Skalen angeht. Im Vordergrund der Analyse steht dabei der Betrag der durch sozialerwünschtes Antwortverhalten zu erwartenden Verzerrung des wahren Vertrauenswertes.

Beide der hier differenzierten Probandengruppen neigen überdurchschnittlich stark zu sozialerwünschtem Antwortverhalten.

7. Hauptstudie

Tab. 7.91: Tendenz zu sozialerwünschtem Antwortverhalten

Gruppe A [Messintention der Objektiv-Skalen erkannt] (N=12)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,6667	5,2500	4,0833
s	1,15470	,86603	1,44338
Gruppe B [Messintention der Objektiv-Skalen nicht erkannt] (N=38)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,6757	4,5405	4,6053
s	,85160	1,19244	1,17495

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*
SDE: *self-deceptive enhancement* (Selbsttäuschung)
IM: *impression management* (Fremdtäuschung)
(Quelle: eigene Berechnungen)

Tab. 7.92: Bivariate Korrelationen: BIDR_(KS) und PR₁ bzw. PT₁ (N = 12)**[Messintention von PR₁ identifiziert]**

	PR ₁	PR _{1(KS)}	PT ₁	BIDR _(KS)	SDE	IM
PR ₁	1	---	---	---	---	---
PR _{1(KS)}	,739**	1	---	---	---	---
PT ₁	,636*	,862**	1	---	---	---
BIDR _(KS)	,455	,092	,273	1	---	---
SDE	-,091	-,492	-,212	,636*	1	---
IM	,455	,222	,273	,782**	,127	1

PR₁: Objektiv-SkalaPR_{1(KS)}: Objektiv-KurzskalaPT₁: Standard-SkalaBIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.93: Bivariate Korrelationen: BIDR_(KS) und PR₁ bzw. PT₁ (N = 37)**[Messintention von PR₁ nicht identifiziert]**

	PR ₁	PR _{1(KS)}	PT ₁	BIDR _(KS)	SDE	IM
PR ₁	1	---	---	---	---	---
PR _{1(KS)}	,840**	1	---	---	---	---
PT ₁	,595**	,591**	1	---	---	---
BIDR _(KS)	,298	,226	,278	1	---	---
SDE	,188	,242	,242	,670**	1	---
IM	,151	,000	,076	,665**	,053	1

PR₁: Objektiv-SkalaPR_{1(KS)}: Objektiv-KurzskalaPT₁: Standard-SkalaBIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Vergleicht man die Korrelationen der Objektiv-Skalen mit der BIDR-Kurzskala in Gruppe A und B, fällt zunächst auf, dass sich der Einfluss sozialerwünschten Antwortverhaltens auf die Objektiv-Skala augenscheinlich in Gruppe B verringert, während er sich für die Objektiv-Kurzskala in Gruppe B verstärkt. Betrachtet man allerdings die Subkomponenten SDE und IM zeigt sich, dass für beide Objektiv-Skalen in Gruppe A tatsächlich ein stärkerer Einfluss sozialer Erwünschtheit auf das Antwortverhalten festzustellen ist als in Gruppe B.

In Gruppe A scheinen sich die aus der Tendenz zur Selbst- und Fremdtäuschung ableitenden Effekte auf das Antwortverhalten wechselseitig zu kompensieren. Zumindest lässt sich dies

anhand der gegenläufigen Vorzeichen der Subkomponenten vermuten. Die Probanden haben somit auf den Objektiv-Skalen in Gruppe A unbewusst aufrichtiger geantwortet.

In Gruppe B neigen die Probanden auf beiden Objektiv-Skalen weniger zu Selbst- und Fremdtäuschung als in Gruppe A. Allerdings weisen die Effekte jeweils die gleichen Wirkrichtungen auf, wie sich an den Vorzeichen der Korrelationen ablesen lässt. Selbst- und Fremdtäuschung verstärken sich somit wechselseitig, was auf der Ebene der BIDR-Kurzskala eine größere Belastung durch sozialerwünschtes Antwortverhalten gegenüber Gruppe A suggeriert. Tatsächlich ist die Belastung gesunken, allerdings sind die Antworten in Gruppe B aufgrund des Verstärkungseffektes stärker verzerrt.

Es kann somit geschlossen werden, dass die Objektiv-Skalen in Gruppe B weniger durch sozialerwünschtes Antwortverhalten beeinflusst sind als in Gruppe A. Jedoch sind die Antworten der Objektiv-Kurzskala in Gruppe B insgesamt stärker verzerrt als in Gruppe A.

Auch für die Standard-Skala lassen sich auf der Ebene der Subkomponenten Unterschiede zwischen Gruppe A und B entdecken. Insgesamt bleibt der Effekt sozialer Erwünschtheit allerdings auf einem ähnlichen Level. Dies war zu erwarten, da die Zieldimension der Standard-Skala grundsätzlich nicht verschleiert war.

Der Zusammenhang zwischen Standard-Skala und SD-Skala bleibt in Gruppe A und B annähernd konstant. Allerdings ist die Standard-Skala in Gruppe B stärker durch sozialerwünschtes Antwortverhalten beeinflusst als in Gruppe A. Die mit der Zieldimension übereinstimmende Augenscheinvalidität der Standard-Skala scheint diese anfällig für sozialerwünschtes Antwortverhalten zu machen.

In Gruppe A ergeben sich erhebliche Unterschiede hinsichtlich der Beeinflussbarkeit der Objektiv-Skalen und der Standard-Skala durch sozialerwünschtes Antwortverhalten. Insgesamt scheint die Standard-Skala am wenigsten durch soziale Erwünschtheit beeinflusst zu sein. Allerdings kann für die Objektiv-Kurzskala angenommen werden, dass ihre Messwerte aufgrund der Kompensation von SDE- und IM-Komponente am geringsten verfälscht sind.

In Gruppe A sind die Objektiv-Skalen und die Standard-Skala unterschiedlich stark bzw. nicht ähnlich stark mit der SD-Skala korreliert. Die Transparenz der Zieldimension bei beiden Skalentypen bewirkt somit nicht, dass sich eine übereinstimmende Anfälligkeit der Skalen für sozialerwünschtes Antwortverhalten ergibt.

In Gruppe B lassen sich keine nennenswerten Unterschiede zwischen dem Einfluss sozialerwünschten Antwortverhaltens auf die Objektiv-Skalen und die Standard-Skala finden. **Es kann daher geschlossen werden, dass innerhalb von Gruppe B die Objektiv-Skalen nicht weniger stark mit der SD-Skala korreliert sind als die Standard-Skala. Vielmehr ist anzunehmen, dass sich kein Unterschied in den Korrelationen ergibt. Entsprechend bewirkt die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen keine Reduktion der Anfälligkeit der Objektiv-Skalen für sozialerwünschtes Antwortverhalten im Vergleich zur Standard-Skala.**

Aus vorherigen Analysen geht hervor, dass sich das Antwortverhalten auf der Standard-Skala und das Antwortverhalten auf den Objektiv-Skalen für die Gruppe der Probanden, die die Messintention der Objektiv-Skalen identifiziert haben, nicht überzufällig voneinander abweicht. Dagegen ergibt sich für die Probanden, die die Messintention nicht erkannt haben, ein

7. Hauptstudie

signifikanter Unterschied im Antwortverhalten auf den genannten Skalen. Nachfolgend soll betrachtet werden, inwieweit die Tendenz zu sozialerwünschtem Antwortverhalten einen Effekt auf die Differenz zwischen den Skalen hat.

Tab. 7.94: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 12)
[Messintention von PR₁ identifiziert (Gruppe A)]

	Diff (PT ₁ - PR ₁)	Diff (PT ₁ - PR _{1(KS)})	BIDR _(KS)	SDE	IM
Diff (PT ₁ - PR ₁)	1	---	---	---	---
Diff (PT ₁ - PR _{1(KS)})	,526	1	---	---	---
BIDR _(KS)	-,039	,348	1	---	---
SDE	-,196	,522	,636*	1	---
IM	-,039	,104	,782**	,127	1

Diff (PT₁ - PR₁): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PT₁ - PR_{1(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.95: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 37)
[Messintention von PR₁ nicht identifiziert (Gruppe B)]

	Diff (PT ₁ - PR ₁)	Diff (PT ₁ - PR _{1(KS)})	BIDR _(KS)	SDE	IM
Diff (PT ₁ - PR ₁)	1	---	---	---	---
Diff (PT ₁ - PR _{1(KS)})	,863**	1	---	---	---
BIDR _(KS)	,085	,120	1	---	---
SDE	,135	,061	,670**	1	---
IM	-,038	,091	,665**	,053	1

Diff (PT₁ - PR₁): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PT₁ - PR_{1(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Beim Vergleich von Gruppe A und B lässt sich feststellen, dass die Differenz zwischen Objektiv-Skala und Standard-Skala in beiden Gruppen nur sehr gering durch sozialerwünschtes Antwortverhalten beeinflusst wird. Die Differenz zwischen Objektiv-Kurzskala und Standard-Skala ist dagegen in beiden Gruppen zumindest in geringem Maße durch sozialerwünschtes Antwortverhalten beeinflusst.

Es lässt sich daher schließen, dass soziale Erwünschtheit zur Erklärung des Abstandes zwischen Objektiv-Skala und Standard-Skala kaum einen Beitrag leistet, während soziale Erwünschtheit für den Abstand zwischen Objektiv-Kurzskala und Standard-Skala als Erklärungsvariable herangezogen werden kann.

In Gruppe A korreliert die Differenz zwischen den Messwerten von Objektiv- und Standard-Skala negativ mit der BIDR-Kurzskala, während die Differenz zwischen den Messwerten von Objektiv-Kurzskala und Standard-Skala positiv mit der BIDR-Kurzskala korreliert ist. Die Richtung des Zusammenhangs setzt sich auch jeweils für die Subkomponenten der BIDR-Kurzskala fort.

Unter der Bedingung, dass die Messintention beider Skalentypen einsichtig war, verringert sich der Unterschied zwischen der Objektiv- und Standard-Skala infolge sozialerwünschten Antwortverhaltens, während der Abstand zwischen Objektiv-Kurzskala und Standard-Skala durch sozialerwünschtes Antwortverhalten vergrößert wird. Aufgrund der

Präsentationsreihenfolge der Skalen ist zu vermuten, dass hierbei die Antworten auf der Standard-Skala an die Antworten auf der Objektiv-Skala angepasst wurden.

In Gruppe B korrelieren die Differenzen zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala auf der Ebene der BIDR-Kurzskala jeweils positiv mit der SD-Skala. Dieser Befund setzt sich auch für die Subkomponenten weitestgehend fort.

Unter der Bedingung, dass die Messintention der Objektiv-Skalen erfolgreich verschleiert wurde, vergrößert sich der Unterschied zwischen den Objektiv-Skalen und der Standard-Skala infolge sozialerwünschten Antwortverhaltens.

In Gruppe A sind die Objektiv-Skalen und die Standard-Skala unterschiedlich stark mit der SD-Skala korreliert, wobei die Vorzeichen dieser Korrelationen einander nicht entgegengesetzt sind. Dabei ist die Objektiv-Skala stärker und die Objektiv-Kurzskala weniger stark durch sozialerwünschtes Antwortverhalten verfälscht als die Standard-Skala. Es ist allerdings zu beachten, dass die geringe Anfälligkeit für Verfälschungen der Objektiv-Kurzskala durch eine wechselseitige Kompensation der SDE- und der IM-Komponente zustande kommt. Auch im Falle der Standard-Skala weisen die Korrelationen mit den Subkomponenten der BIDR-Kurzskala entgegengesetzte Vorzeichen auf, aus denen sich jedoch kein Kompensationseffekt ergibt. Außerdem ist die BIDR-Kurzskala nur mit der Differenz von Objektiv- und Standard-Skala negativ korreliert.

Die Angleichung der Messwerte von Objektiv- und Standard-Skala scheint hier vornehmlich auf das Antwortverhalten auf der Objektiv-Skala zurückzugehen, während die Vergrößerung des Abstandes der Messwerte von Objektiv-Kurzskala und Standard-Skala wohl auf das Antwortverhalten auf der Standard-Skala zurückzuführen ist. Die Probanden überschätzen ihr Vertrauen auf allen Vertrauensskalen, allerdings überschätzen sie es auf der Objektiv-Skala stärker als auf der Standard-Skala und auf der Objektiv-Kurzskala weniger stark als auf der Standard-Skala.

In Gruppe B sind die Objektiv-Skalen und die Standard-Skala ähnlich stark mit der SD-Skala korreliert. Allerdings ist die Objektiv-Skala etwas stärker und die Objektiv-Kurzskala etwas weniger stark durch sozialerwünschtes Antwortverhalten verfälscht als die Standard-Skala. Die Differenzen zwischen den Objektiv-Skalen und der Standard-Skala sind dabei jeweils positiv mit der BIDR-Kurzskala korreliert.

Die Vergrößerung des Unterschiedes zwischen den Messwerten der Objektiv-Skala und der Standard-Skala scheint hierbei vornehmlich dem Antwortverhalten auf der Objektiv-Skala anzulasten zu sein, während die Abstandsvergrößerung zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala vornehmlich auf das Antwortverhalten auf der Standard-Skala zurückgeht. Die Probanden überschätzen ihr Vertrauen auf allen Vertrauensskalen, allerdings überschätzen sie es auf der Objektiv-Skala stärker als auf der Standard-Skala und auf der Objektiv-Kurzskala weniger stark als auf der Standard-Skala.

In der Zusammenschau bedeutet dies, dass zumindest der Unterschied zwischen Objektiv-Kurzskala und Standard-Skala dadurch zustande kommt, dass die Probanden ihr Vertrauen auf der Standard-Skala überschätzen, während sie auf der Objektiv-Kurzskala tendenziell aufrichtiger antworten. Dieser Effekt ist allerdings unabhängig von der Verschleierung der Messintention. Die Objektiv-Skala trägt dagegen nicht zur Verbesserung der Messgüte bei, da

7. Hauptstudie

sie sowohl in Gruppe A als auch Gruppe B für den Unterschied zwischen den Messwerten von Objektiv- und Standard-Skala verantwortlich zu sein scheint.

7.2.7.1.2 Differenzierung nach Identifikation der Messintention und Referenzgruppe

Berücksichtigt man das Radiusproblems der Vertrauensmessung, so wird mit Hilfe der Vertrauensskalen strenggenommen nur dann Generalisiertes Vertrauen erfasst, wenn die Probanden generalisierte Andere bzw. die *outgroup* als Vertrauensreferenz angeben. Diesem Umstand soll auch im Rahmen der Analyse des Einflusses sozialer Erwünschtheit Rechnung getragen werden. Nachfolgend werden vorausgegangene Analyseschritte anhand von Samples wiederholt, in die nur Probanden aus *Versuchsgruppe 1* eingehen, welche die *outgroup* als Bezugspunkt ihres Vertrauens angegeben haben. Außerdem wird danach differenziert, ob die Zieldimension der Objektiv-Skalen aufgedeckt werden konnte oder nicht. Dadurch wird die in die Zusammenhangsanalyse eingehende Fallzahl allerdings deutlich reduziert.

Tab. 7.96: Tendenz zu sozialerwünschtem Antwortverhalten (Vertrauensreferenz entspricht *outgroup*)

Gruppe A [Messintention der Objektiv-Skalen erkannt] (N=8)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,3750	5,1250	3,7500
<i>s</i>	1,18773	,99103	1,38873
Gruppe B [Messintention der Objektiv-Skalen <u>nicht</u> erkannt] (N=14)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,7143	4,5000	4,5714
<i>s</i>	,82542	1,01905	1,15787

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

(Quelle: eigene Berechnungen)

Auf der Ebene der BIDR-Kurzskala fällt die Neigung zu sozialerwünschtem Antwortverhalten für die Gruppen A und B nahezu gleichgroß aus. Beide Gruppen neigen insgesamt eher dazu, sich positiver darzustellen als es der Wirklichkeit entspricht. Für die Subkomponenten lassen sich dagegen Unterschiede finden, so ist für Gruppe A die Neigung zur Selbsttäuschung stärker ausgeprägt als in Gruppe B, während für Fremdtäuschung die umgekehrte Relation gilt.

Tab. 7.97: Bivariate Korrelationen: BIDR_(KS) und PR₁ bzw. PT₁ (N = 8)

[Messintention von PR₁ identifiziert und Vertrauensreferenz entspricht *outgroup* (Gruppe A)]

	PR ₁	PR _{1(KS)}	PT ₁	BIDR _(KS)	SDE	IM
PR ₁	1	---	---	---	---	---
PR _{1(KS)}	,832*	1	---	---	---	---
PT ₁	,816*	,793*	1	---	---	---
BIDR _(KS)	,195	,090	,557	1	---	---
SDE	-,234	-,540	-,095	,683	1	---
IM	,333	,401	,680	,758*	,130	1

PR₁: Objektiv-Skala

PR_{1(KS)}: Objektiv-Kurzskala

PT₁: Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.98: Bivariate Korrelationen: BIDR_(KS) und PR₁ bzw. PT₁ (N = 14)

[Messintention von PR₁ nicht identifiziert und Vertrauensreferenz entspricht *outgroup* (Gruppe B)]

	PR ₁	PR _{1(KS)}	PT ₁	BIDR _(KS)	SDE	IM
PR ₁	1	---	---	---	---	---
PR _{1(KS)}	,707**	1	---	---	---	---
PT ₁	,347	,420	1	---	---	---
BIDR _(KS)	,533*	,252	,357	1	---	---
SDE	,557*	,557*	,536*	,640*	1	---
IM	,220	-,220	,015	,586*	-,065	1

PR₁: Objektiv-Skala

PR_{1(KS)}: Objektiv-Kurzskala

PT₁: Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Ein Vergleich der Korrelationen der Objektiv-Skalen mit der BIDR-Kurzskala zwischen den Gruppen A und B zeigt zunächst, dass sich der Einfluss sozialerwünschten Antwortverhaltens auf die Objektiv-Skalen in Gruppe B verstärkt. Betrachtet man dagegen die Subkomponenten SDE und IM lässt sich zumindest für die Objektiv-Kurzskala feststellen, dass tatsächlich ein leichter Rückgang der Belastung durch soziale Erwünschtheit (s. IM-Komponente) in Gruppe B zu verzeichnen ist. Für die Objektiv-Skala hat dagegen erstgenannter Befund weiterhin Bestand.

In Gruppe A scheinen sich die aus der Tendenz zur Selbst- und Fremdtäuschung resultierenden Effekte auf das Antwortverhalten bei beiden Objektiv-Skalen wechselseitig zu kompensieren. So sind die Vorzeichen der Subkomponenten jeweils gegenläufig. In Summe haben die Probanden somit in Gruppe A auf den Objektiv-Skalen unbewusst aufrichtiger geantwortet, wobei das Antwortverhalten tatsächlich stark durch soziale Erwünschtheit beeinflusst ist.

In Gruppe B ist die Objektiv-Skala stark durch sozialerwünschtes Antwortverhalten belastet. Die Selbsttäuschungs- und die Fremdtäuschungskomponente verstärken sich hier wechselseitig, was eine erhebliche Verzerrung der Antworten zur Folge hat. Für die Objektiv-Kurzskala lässt sich feststellen, dass sich der Einfluss von Fremdtäuschung auf das Antwortverhalten im Vergleich zu Gruppe A etwas reduziert. Der Einfluss der Selbsttäuschungskomponente bleibt dagegen auf einem ähnlichen Level. Wie in Gruppe A kompensieren sich auch in Gruppe B bei der Objektiv-Kurzskala die Effekte der SDE- und der IM-Komponente zumindest in Teilen. Die Probanden haben auf der Objektiv-Kurzskala in Gruppe B somit unbewusst aufrichtiger geantwortet, was allerdings auf die unterschiedlichen Wirkrichtungen der Subkomponenten zurückzuführen ist.

Es kann somit geschlossen werden, dass zumindest die Objektiv-Kurzskala in Gruppe B weniger stark durch soziale Erwünschtheit beeinflusst ist. Allerdings sind ihre Messwerte infolge der Kompensation von SDE- und IM-Tendenz in Gruppe A weniger durch sozialerwünschtes Antwortverhalten verzerrt als in Gruppe B. Für die Objektiv-Skala lässt sich dagegen feststellen, dass sie in Gruppe A weniger stark mit der SD-Skala korreliert ist als in Gruppe B. Die erfolgreiche Verschleierung der Messintention hat im Fall der Objektiv-Kurzskala somit eine Reduktion der Anfälligkeit für sozialerwünschtes Antwortverhalten zur Folge, die sich allerdings nicht in den Messergebnissen niederschlägt.

Für die Standard-Skala lässt sich bei Betrachtung der Subkomponenten der BIDR-Kurzskala feststellen, dass sich zwischen Gruppe A und B keine nachhaltige Veränderung ergibt, was die Belastung durch sozialer Erwünschtheit angeht. Allerdings spielt in Gruppe A vornehmlich die Tendenz zur Fremdtäuschung eine Rolle, während in Gruppe B vornehmlich die Tendenz zur Selbsttäuschung für die hohe Belastung verantwortlich ist.

Die Korrelation der Messwerte der Standard-Skala mit der SD-Skala fällt in Gruppe A und B vergleichsweise ähnlich hoch aus. Die Transparenz der Messintention der Standard-Skala scheint für diese Anfälligkeit für sozialerwünschtes Antwortverhalten verantwortlich zu sein.

In Gruppe A ergeben sich zwischen den Objektiv-Skalen und der Standard-Skala vordergründig deutliche Unterschiede hinsichtlich der Anfälligkeit für sozialerwünschtes Antwortverhalten. Betrachtet man jedoch die Subkomponenten der BIDR-Kurzskala zeigt sich, dass die Unterschiede weniger stark ausfallen. Allerdings bewirkt die Kompensation von SDE- und IM-Komponente für die Objektiv-Skalen, dass deren Messwerte insgesamt weniger verfälscht sind.

Sind die Zieldimension bei beiden Skalentypen transparent, so sind die Messwerte der Standard-Skala stärker durch sozialerwünschtes Antwortverhalten verzerrt als die Messwerte der Objektiv-Skalen. Dieser Effekt wird allerdings durch eine Kompensation der Selbst- und Fremdtäuschungstendenz bei den Objektiv-Skalen bewirkt. Insbesondere die Objektiv-Kurzskala scheint dadurch jedoch stärker durch soziale Erwünschtheit beeinflusst zu sein als die Standard-Skala.

In Gruppe B haben die Probanden auf der Objektiv-Kurzskala zwar insgesamt aufrichtiger geantwortet als auf der Standard-Skala, allerdings ist die Belastung der Objektiv-Kurzskala durch sozialerwünschtes Antwortverhalten größer, wie sich an den Subkomponenten der BIDR-Kurzskala ablesen lässt. Für die Objektiv-Skala ist die Belastung durch sozialerwünschtes Antwortverhalten höher als für die Standard-Skala und auch die Messwerte der Objektiv-Skala sind stärker durch sozialerwünschtes Antwortverhalten verzerrt als die der Standard-Skala.

Es kann daher geschlossen werden, dass innerhalb von Gruppe B nur die Messwerte der Objektiv-Kurzskala nicht weniger durch sozialerwünschtes Antwortverhalten verfälscht werden als die der Standard-Skala. Daneben scheint allerdings auch zu gelten, dass die Objektiv-Skalen stärker durch soziale Erwünschtheit beeinflusst sind als die Standard-Skala. Die schlägt sich allerdings nur in den Resultaten der Objektiv-Skala nieder. Die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen hat somit nur im Falle der Objektiv-Kurzskala einen reduzierenden Effekt auf die Verfälschbarkeit durch sozialerwünschtes Antwortverhalten.

Nachfolgend soll geprüft werden, inwiefern die Tendenz zu sozialerwünschtem Antwortverhalten mit der potentiellen Differenz zwischen den durchschnittlichen Summenwerten der Objektiv-Skalen und der Standard-Skala zusammenhängen. Je nach Skalenpaarung und Identifikation der Messintention ergaben die zugehörigen Signifikanzprüfungen unterschiedliche Ergebnisse. Dabei konnte keine klare Tendenz hinsichtlich des Einflusses der geringen Augenscheinvalidität der Objektiv-Items festgestellt werden. Auch hinsichtlich dieser Analyse gilt es zu beachten, dass die eingehende Fallzahl stark reduziert ist.

Tab. 7.99: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 8)
[Messintention von PR₁ identifiziert und Vertrauensreferenz entspricht *outgroup*
(Gruppe A)]

	Diff (PT ₁ - PR ₁)	Diff (PT ₁ - PR _{1(KS)})	BIDR _(KS)	SDE	IM
Diff (PT ₁ - PR ₁)	1	---	---	---	---
Diff (PT ₁ - PR _{1(KS)})	,447	1	---	---	---
BIDR _(KS)	,715*	,668	1	---	---
SDE	,078	,731*	,683	1	---
IM	,778*	,348	,758*	,130	1

Diff (PT₁ - PR₁): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PT₁ - PR_{1(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.100: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 14)
[Messintention von PR₁ nicht identifiziert und Vertrauensreferenz entspricht *outgroup*
(Gruppe B)]

	Diff (PT ₁ - PR ₁)	Diff (PT ₁ - PR _{1(KS)})	BIDR _(KS)	SDE	IM
Diff (PT ₁ - PR ₁)	1	---	---	---	---
Diff (PT ₁ - PR _{1(KS)})	,835**	1	---	---	---
BIDR _(KS)	-,041	,174	1	---	---
SDE	,117	,123	,640*	1	---
IM	-,147	,186	,586*	-,065	1

Diff (PT₁ - PR₁): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PT₁ - PR_{1(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

In Gruppe A korreliert die Differenzen zwischen den Objektiv-Skalen und der Standard-Skala jeweils stark mit der BIDR-Kurzskala. In Gruppe B besteht zwischen der Differenz von Objektiv- und Standard-Skala kein Zusammenhang mit der BIDR-Kurzskala, wohingegen die Differenz zwischen Objektiv-Kurzskala und Standard-Skala mit der BIDR-Kurzskala korreliert ist.

Entsprechend lässt sich ableiten, dass in Gruppe A sozialerwünschtes Antwortverhalten einen erheblichen Einfluss auf die Differenzen zwischen den Objektiv-Skalen und der Standard-Skala hat. In Gruppe B kann soziale Erwünschtheit allerdings nur für die Differenz zwischen Objektiv-Kurzskala und Standard-Skala als Erklärungsvariable herangezogen werden.

In Gruppe A sind die Differenzen zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala jeweils positiv mit der BIDR-Kurzskala korreliert. Diese Richtung des Zusammenhangs gilt jeweils auch für die Subkomponenten der BIDR-Kurzskala.

Unter der Bedingung, dass die Messintention beider Skalentypen transparent war, vergrößert sich der Unterschied zwischen den Objektiv-Skalen und der Standard-Skala infolge sozialerwünschten Antwortverhaltens.

In Gruppe B korreliert die Differenz zwischen den Messwerten von Objektiv- und Standard-Skala negativ mit der BIDR-Kurzskala, während die Differenz zwischen den Messwerten von

Objektiv-Kurzskala und Standard-Skala positiv mit der BIDR-Kurzskala korreliert ist. Diese Richtung des Zusammenhangs setzt sich für die Subkomponenten allerdings bei der Differenz von Objektiv-Kurzskala und Standard-Skala fort. Dagegen lässt sich bei der Differenz von Objektiv- und Standard-Skala eine wechselseitige Kompensation der SDE- und der IM-Komponente feststellen.

Der Unterschied zwischen Objektiv- und Standard-Skala verringert sich demnach infolge sozialerwünschten Antwortverhaltens, während der Abstand zwischen Objektiv-Kurzskala und Standard-Skala durch sozialerwünschtes Antwortverhalten vergrößert wird. Bei erfolgreicher Verschleierung der Messintention ergibt sich nur im Falle der Objektiv-Kurzskala eine Vergrößerung der Messwertdifferenz infolge sozialerwünschten Antwortverhaltens. Dagegen gleichen sich die Messwerte von Objektiv- und Standard-Skala aneinander an. Aufgrund der Präsentationsreihenfolge der Skalen ist zu vermuten, dass hierbei die Antworten auf der Standard-Skala an die Antworten auf der Objektiv-Skala angepasst wurden.

In Gruppe A sind die Objektiv-Skalen und die Standard-Skala auf der Ebene der BIDR-Kurzskala unterschiedlich stark mit der SD-Skala korreliert und weisen keine einander entgegengesetzte Vorzeichen auf. Dabei sind die Objektiv-Skalen jeweils weniger stark durch sozialerwünschtes Antwortverhalten verfälscht als die Standard-Skala. Hierbei ist jedoch zu beachten, dass dieser Effekt durch eine wechselseitige Kompensation des Einflusses von Selbsttäuschungs- und Fremdtäuschungstendenz zustande kommt. Die BIDR-Kurzskala ist jeweils positiv mit den Differenzen zwischen den Skalentypen korreliert.

Die Abstandsvergrößerung zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala scheint hier vornehmlich auf das Antwortverhalten auf der Standard-Skala zurückzugehen. Die Probanden überschätzen zwar ihr wahres Vertrauen auf allen Vertrauensskalen, allerdings fällt die Überschätzung auf der Standard-Skala höher aus als auf den Objektiv-Skalen.

In Gruppe B sind die Objektiv-Skalen und die Standard-Skala auf der Ebene der BIDR-Kurzskala unterschiedlich stark mit der SD-Skala korreliert. Dabei ist die Objektiv-Skala stärker und die Objektiv-Kurzskala weniger stark durch sozialerwünschtes Antwortverhalten verfälscht als die Standard-Skala. Die Differenz zwischen Objektiv- und Standard-Skala ist negativ und die Differenz zwischen Objektiv-Kurzskala und Standard-Skala ist positiv mit der BIDR-Kurzskala korreliert.

Die Abstandsverringerung zwischen den Messwerten der Objektiv- und Standard-Skala scheinen vornehmlich auf das Antwortverhalten auf der Objektiv-Skala zurückzugehen. Die Probanden überschätzen ihr wahres Vertrauen auf der Objektiv-Skala stärker als auf der Standard-Skala. Die Abstandsvergrößerung zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala ist indes vornehmlich auf das Antwortverhalten auf der Standard-Skala zurückzuführen. Hier überschätzen die Probanden ihr wahres Vertrauen auf der Standard-Skala, während sie auf der Objektiv-Kurzskala tendenziell aufrichtiger antworten.

In der Zusammenschau bedeutet dies, dass zumindest der Unterschied zwischen Objektiv-Kurzskala und Standard-Skala dadurch zustande kommt, dass die Probanden ihr Vertrauen auf der Standard-Skala überschätzen, während sie auf der Objektiv-Kurzskala tendenziell aufrichtiger antworten. Dieser Effekt ist allerdings unabhängig von der Verschleierung der Messintention. Die Verschleierung der Messintention trägt im Falle der Objektiv-Skala nicht

zur Verbesserung der Messgüte bei, da sie in Gruppe B für den Unterschied zwischen den Messwerten von Objektiv- und Standard-Skala verantwortlich zu sein scheint.

7.2.7.2 Gruppe 2: Standard/Objektiv

Auch die Probanden in *Versuchsgruppe 2* besitzen eine überdurchschnittliche Neigung sozialerwünscht zu antworten und weisen dabei eine leicht geringere Varianz auf als die Probanden in *Versuchsgruppe 1* ($\bar{x}_{BIDR} = 4,68$, $s = 0,819$; $\bar{x}_{SDE} = 4,82$, $s = 0,941$; $\bar{x}_{IM} = 4,56$, $s = 1,215$). Aus der grafischen Analyse geht hervor, dass insbesondere die Verteilungen für die vollständige SD-Skala und für die SDE-Komponente deutlich rechtssteil bzw. linksschief sind und demnach eine starke Tendenz zu sozialerwünschtem Antwortverhalten besteht. Auch bezüglich der IM-Komponente ordnet sich die Mehrzahl der Befragten auf der rechten Hälfte der Antwortskala ein und zeigt dadurch eine erhöhte Neigung zu Fremdtäuschung an.

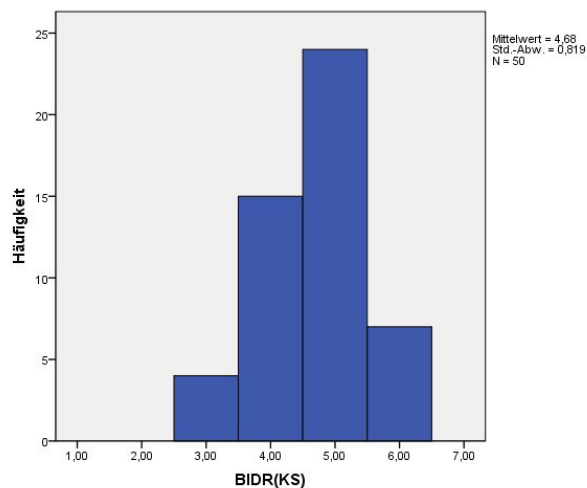


Abb. 7.28: Häufigkeitsverteilung Summenscore BIDR-Kurzska („soziale Erwünschtheit“) (mit 1 = geringe Neigung zu SD bis 5 = starke Neigung zu SD) in Versuchsgruppe 2

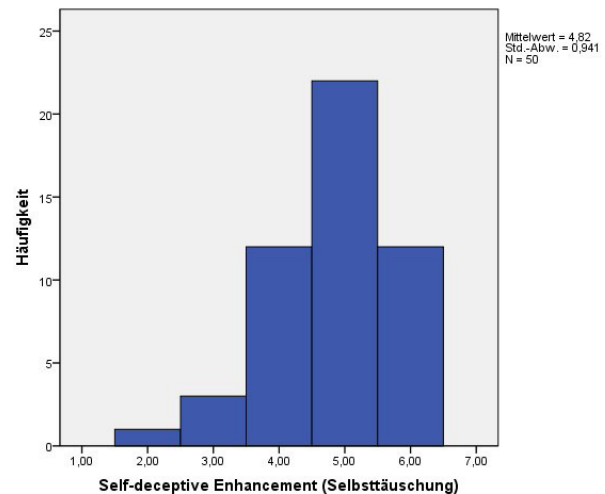


Abb. 7.29: Häufigkeitsverteilung Summenscore SDE-Komponente („Selbsttäuschung“) (mit 1 = geringe Neigung zu SDE bis 5 = starke Neigung zu SDE) in Versuchsgruppe 2

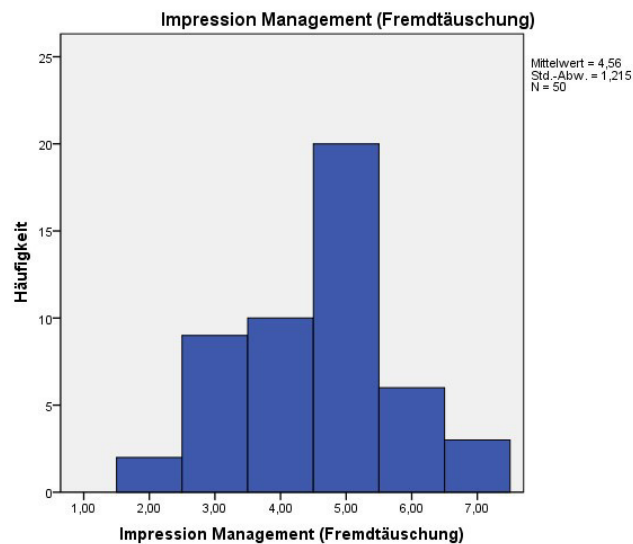


Abb. 7.30: Häufigkeitsverteilung Summenscore IM-Komponente („Fremdtäuschung“) (mit 1 = geringe Neigung zu IM bis 5 = starke Neigung zu IM) in Versuchsgruppe 2

Tab. 7.101: Bivariate Korrelationen: BIDR_(KS) und PR₂ bzw. PT₂ (N = 50)

	PR ₂	PT ₂	PT _{2(KS)}	BIDR _(KS)	SDE	IM
PR ₂	1	---	---	---	---	---
PT ₂	,502**	1	---	---	---	---
PT _{2(KS)}	,603**	,640**	1	---	---	---
BIDR _(KS)	-,017	-,032	,011	1	---	---
SDE	-,161	-,148	-,233	,559**	1	---
IM	,094	,004	,160	,737**	,036	1

PR₂: Standard-Skala

PT₂: Objektiv-Skala

PT_{2(KS)}: Objektiv-Kurzskala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Betrachtet man die Korrelationen der Vertrauensskalen mit der BIDR-Kurzskala, zeigt sich, dass keine der Skalen nennenswert durch sozialerwünschtes Antwortverhalten beeinflusst ist.

Analysiert man die Subkomponenten, ergibt sich ein etwas anders gelagertes Bild. Hier sind zumindest Zusammenhänge von geringer Stärke zu beobachten. Insbesondere eine Tendenz zur Selbsttäuschung scheint bei der Bearbeitung der Skalen eine Rolle gespielt zu haben, allerdings entgegengesetzt der erwarteten Richtung. So scheint sich die Vertrauensneigung zu steigern, wenn die Tendenz zur Selbsttäuschung abnimmt und umgekehrt. Die Probanden neigen somit im Falle eines in optimistischer Weise verzerrten Blickwinkels auf die Realität dazu, ein weniger entwickeltes Generalisiertes Vertrauen aufzuweisen. Dieser Befund spricht allerdings gegen die theoretischen Annahmen der Vertrauenstheorie, wonach eine durch Optimismus geprägte Weltanschauung ein erhöhtes Generalisiertes Vertrauen nach sich zieht. Eine alternative Erklärung für dieses Ergebnis lässt sich aus der Interpretation von Selbsttäuschung als Schutz des Selbstbildes und des Selbstwertgefühls ableiten. Möglicherweise stellt für die Probanden ein gesteigertes Generalisiertes Vertrauen, d.h. eine ausgeprägte undifferenzierte

7. Hauptstudie

Vertrauensdisposition, ein Ausdruck von Naivität dar. Da Naivität ein eher negativ besetztes Charaktermerkmal repräsentiert, ergibt sich der negative Zusammenhang von Selbsttäuschung und Generalisiertem Vertrauen vermutlich deshalb, weil die Probanden vermeiden möchten, als zu vertrauensselig zu erscheinen. Entsprechend geht die Bewahrung eines positiven Selbstbildes mit einem Absinken des Generalisierten Vertrauens einher.

Fremdtäuschung scheint nur für das Antwortverhalten auf der Objektiv-Kurzskala eine gewisse Relevanz zu besitzen. So geht hier eine Tendenz zur Fremdtäuschung mit einem Anstieg des Generalisierten Vertrauens einher. Die Tendenz zur Selbsttäuschung wirkt diesem Trend entgegen, sodass die Probanden auf der Objektiv-Kurzskala in der Summe verhältnismäßig ehrlich antworten. Mit anderen Worten: Die Objektiv-Kurzskala ist zwar vergleichsweise stark durch sozialerwünschtes Antwortverhalten belastet, es scheint sich dadurch allerdings keine Konsequenz für die Güte der Messung zu ergeben.

In der Zusammenschau wird innerhalb von *Versuchsgruppe 2* weder durch die Objektiv-Skala noch durch die Objektiv-Kurzskala eine im Vergleich zur Standard-Skala gesteigerte Validität der Vertrauensmessung bewirkt. Während die Objektiv-Skala hinsichtlich der Beeinflussbarkeit durch sozialerwünschtes Antwortverhalten ähnliche Parameter aufweist wie die Standard-Skala, wird das Antwortverhalten auf der Objektiv-Kurzskala vergleichsweise stark durch soziale Erwünschtheit beeinträchtigt.

Es lässt sich daher schließen, dass die Objektiv-Skalen nicht weniger stark mit der SD-Skala korreliert sind als die Standard-Skala. Vielmehr scheint sich kein nachhaltiger Unterschied zu ergeben, was die Belastung durch sozialerwünschtes Antwortverhalten angeht. Die Objektiv-Skalen sind somit nicht weniger anfällig für sozialerwünschtes Antwortverhalten als die Standard-Skala.

Auch im Rahmen von *Versuchsgruppe 2* wurden Mittelwertvergleiche bezüglich der Summenscores der Objektiv-Skalen und der Standard-Skala durchgeführt, aus denen hervorging, dass sich die Objektiv-Skalen signifikant von der Standard-Skala unterscheiden. Wie schon in *Versuchsgruppe 1* gaben auch hier die Probanden auf den Objektiv-Skalen ein geringeres Generalisiertes Vertrauen als auf der Standard-Skala an. Im Folgenden soll untersucht werden welchen Einfluss soziale Erwünschtheit auf diese Differenz ausübt.

Tab. 7.102: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 50)

	Diff (PR ₂ - PT ₂)	Diff (PR ₂ - PT _{2(KS)})	BIDR _(KS)	SDE	IM
Diff (PR ₂ - PT ₂)	1	---	---	---	---
Diff (PR ₂ - PT _{2(KS)})	,789**	1	---	---	---
BIDR _(KS)	,003	-,031	1	---	---
SDE	-,080	-,007	,559**	1	---
IM	,105	-,016	,737**	,036	1

Diff (PR₂ - PT₂): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PR₂ - PT_{2(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Die BIDR-Kurzskala korreliert weder mit der Differenz von Objektiv- und Standard-Skala noch mit der Differenz von Objektiv-Kurzskala und Standard-Skala nennenswert.

Hieraus lässt sich ableiten, dass soziale Erwünschtheit keinen Beitrag zur Aufklärung des Unterschiedes zwischen den Objektiv-Skalen und der Standard-Skala leistet.

7.2.7.2.1 Differenzierung nach Identifikation der Messintention

Nachfolgend soll nun auch für *Versuchsgruppe 2* untersucht werden, inwiefern sich der Einfluss sozialer Erwünschtheit auf die verschiedenen Vertrauensskalen verändert, je nachdem, ob die Messintention der Objektiv-Skalen durch die Probanden identifiziert wurde oder nicht. Auch hier steht wie bereits in *Versuchsgruppe 1* der Betrag der durch sozialerwünschtes Antwortverhalten zu erwartenden Verzerrung des wahren Vertrauenswertes im Fokus der Analyse.

Tab. 7.103: Tendenz zu sozialerwünschtem Antwortverhalten

Gruppe A [Messintention der Objektiv-Skalen erkannt] (N=24)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,6250	4,7500	4,5417
<i>s</i>	,76967	,98907	1,35066
Gruppe B [Messintention der Objektiv-Skalen nicht erkannt] (N=26)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,7308	4,8846	4,5769
<i>s</i>	,87442	,90893	1,10175

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

(Quelle: eigene Berechnungen)

Auch in der zweiten Versuchsgruppe neigen die nach Identifikation der Messintention der Objektiv-Skalen differenzierten Probandengruppen gleichermaßen überdurchschnittlich zu sozialerwünschtem Antwortverhalten.

Tab. 7.104: Bivariate Korrelationen: BIDR_(KS) und PR₂ bzw. PT₂ (N = 24)

[Messintention von PT₂ identifiziert (Gruppe A)]

	PR ₂	PT ₂	PT _{2(KS)}	BIDR _(KS)	SDE	IM
PR ₂	1	---	---	---	---	---
PT ₂	,356	1	---	---	---	---
PT _{2(KS)}	,669**	,540**	1	---	---	---
BIDR _(KS)	,000	-,142	-,218	1	---	---
SDE	-,269	-,074	-,414*	,557**	1	---
IM	,145	-,314	,051	,664**	-,122	1

PR₂: Standard-Skala

PT₂: Objektiv-Skala

PT_{2(KS)}: Objektiv-Kurzskala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.105: Bivariate Korrelationen: BIDR_(KS) und PR₂ bzw. PT₂ (N = 26)
[Messintention von PT₂ nicht identifiziert (Gruppe B)]

	PR ₂	PT ₂	PT _{2(KS)}	BIDR _(KS)	SDE	IM
PR ₂	1	---	---	---	---	---
PT ₂	,633**	1	---	---	---	---
PT _{2(KS)}	,538**	,742**	1	---	---	---
BIDR _(KS)	-,033	,023	,218	1	---	---
SDE	-,053	-,241	-,038	,563**	1	---
IM	,037	,304	,304	,832**	,229	1

PR₂: Standard-Skala

PT₂: Objektiv-Skala

PT_{2(KS)}: Objektiv-Kurzskala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Der Vergleich von Gruppe A und B zeigt, dass die Objektiv-Skala auf der Ebene der BIDR-Kurzskala in Gruppe B weniger durch sozialerwünschtes Antwortverhalten beeinflusst zu sein scheint als in Gruppe A. Dagegen ergeben sich für die Objektiv-Kurzskala keine Unterschiede. Betrachtet man allerdings die Subkomponenten der BIDR-Skala, fällt auf, dass sich für die Objektiv-Skala in Gruppe B tatsächlich ein stärkerer Effekt sozialer Erwünschtheit ergibt als in Gruppe A.

In Gruppe A kommt die vermeintlich höhere Belastung durch sozialerwünschtes Antwortverhalten dadurch zustande, dass im Falle der Objektiv-Skala fast ausschließlich eine Tendenz zur Fremdtäuschung besteht, während im Falle der Objektiv-Kurzskala fast ausschließlich eine Neigung zur Selbsttäuschung beobachtet werden kann.

In Gruppe B kommt die vermeintlich niedrigere Belastung der Objektiv-Skala dadurch zustande, dass sich der Einfluss der Selbsttäuschungskomponente im Vergleich zu Gruppe A verstärkt, während der Einfluss der Fremdtäuschungskomponente gleich groß bleibt. In Gruppe B kompensieren sich beide Einflüsse aufgrund gegenläufiger Vorzeichen. Dies hat zur Folge, dass die Probanden auf der Objektiv-Skala zwar unbewusst aufrichtiger antworten, diese Aufrichtigkeit allerdings gegenläufigen Antwortverzerrungen entspringt. Für die Objektiv-Kurzskala spielt dagegen in Gruppe B ausschließlich die Neigung zur Fremdtäuschung eine Rolle.

Die Messwerte der Objektiv-Skala sind in Gruppe B weniger stark durch sozialerwünschtes Antwortverhalten verzerrt als in Gruppe A. Allerdings ist die Objektiv-Skala in Gruppe B stärker durch soziale Erwünschtheit beeinflusst als in Gruppe A. Die Messwerte der Objektiv-Kurzskala sind in Gruppe A und B ähnlich stark durch sozialerwünschtes Antwortverhalten verzerrt. In Gruppe B geht allerdings der Einfluss sozialer Erwünschtheit im Vergleich zu Gruppe A leicht zurück. Die Verschleierung der Messintention bewirkt somit nur bei der Objektiv-Skala eine Reduktion der Verfälschbarkeit der Messwerte.

Bei Betrachtung der Subkomponenten ist die Standard-Skala in Gruppe B deutlich weniger durch sozialerwünschtes Antwortverhalten beeinflusst als in Gruppe A. Allerdings sind deren Messwerte aufgrund der wechselseitigen Kompensation von Selbst- und Fremdtäuschung in Gruppe A weniger stark durch soziale Erwünschtheit verfälscht.

Die Transparenz der Messintention der Standard-Skala scheint an dieser Stelle somit eher keinen Effekt auf deren Anfälligkeit für sozialerwünschtes Antwortverhalten zu haben.

In Gruppe A ergeben sich auf den ersten Blick erhebliche Unterschiede zwischen den Objektiv-Skalen und der Standard-Skala hinsichtlich deren Beeinflussbarkeit durch sozialerwünschtes Antworten. Betrachtet man die Subkomponenten der BIDR-Skala, relativiert sich dieser Eindruck. Die verschiedenen Skalentypen scheinen jeweils ähnlich stark durch soziale Erwünschtheit beeinflusst zu sein. Bei den Objektiv-Skalen sind entweder eine stark ausgeprägte Fremdtäuschungs- oder Selbsttäuschungstendenz für diesen Effekt verantwortlich. Für die Standard-Skala ergibt sich der oben bereits angesprochene Kompensationseffekt.

Sind die Zieldimensionen bei beiden Skalentypen transparent, werden die Messwerte der Objektiv-Skalen stärker durch sozialerwünschtes Antwortverhalten verfälscht als die der Standard-Skala. Jedoch scheinen beide Skalentypen ähnlich stark durch soziale Erwünschtheit belastet zu sein.

Die Messwerte der Objektiv-Kurzskaala sind in Gruppe B ähnlich stark durch sozialerwünschtes Antwortverhalten verfälscht wie die Messwerte der Standard-Skala. Dagegen sind die Messwerte der Objektiv-Skala stärker verfälscht als die der Standard-Skala. Beide Objektiv-Skalen sind jedoch stärker durch soziale Erwünschtheit beeinflusst als die Standard-Skala.

Die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen macht diese somit im Vergleich mit der Standard-Skala nicht weniger anfällig für sozialerwünschtes Antwortverhalten.

Im Rahmen der Untersuchung von Unterschiedshypothesen zeigte sich, dass sich zwischen dem Antwortverhalten auf der Standard-Skala und dem Antwortverhalten auf der Objektiv-Skala unabhängig davon, ob die Messintention der Objektiv-Skalen aufgedeckt wurde, ein signifikanter Unterschied ergibt. Dagegen zeigt sich zwischen Standard-Skala und Objektiv-Kurzskaala in keiner der hier differenzierten Probandengruppen ein überzufälliger Unterschied im Antwortverhalten. Nachfolgend soll nun untersucht werden, welchen Effekt soziale Erwünschtheit auf die Differenz zwischen den verschiedenen Vertrauensskalen ausübt.

Tab. 7.106: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 24)
[Messintention von PT₂ identifiziert (Gruppe A)]

	Diff (PR ₂ - PT ₂)	Diff (PR ₂ - PT _{2(KS)})	BIDR _(KS)	SDE	IM
Diff (PR ₂ - PT ₂)	1	---	---	---	---
Diff (PR ₂ - PT _{2(KS)})	,757**	1	---	---	---
BIDR _(KS)	,082	,209	1	---	---
SDE	-,240	,036	,557**	1	---
IM	,332	,146	,664**	-,122	1

Diff (PR₂ - PT₂): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PR₂ - PT_{2(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskaala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.107: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 24)
[Messintention von PT₂ nicht identifiziert (Gruppe B)]

	Diff (PR ₂ - PT ₂)	Diff (PR ₂ - PT _{2(KS)})	BIDR _(KS)	SDE	IM
Diff (PR ₂ - PT ₂)	1	---	---	---	---
Diff (PR ₂ - PT _{2(KS)})	,842**	1	---	---	---
BIDR _(KS)	-,063	-,198	1	---	---
SDE	,137	-,035	,563**	1	---
IM	-,213	-,178	,832**	,229	1

Diff (PR₂ - PT₂): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PR₂ - PT_{2(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Aus dem Vergleich der Gruppen A und B geht hervor, dass die Differenz zwischen Objektiv- und Standard-Skala in keiner der beiden Gruppen nennenswert mit der BIDR-Kurzskala korreliert. Die Differenz zwischen Objektiv-Kurzskala und Standard-Skala ist dagegen in beiden Gruppen mit der BIDR-Kurzskala korreliert.

Das bedeutet: Sozialerwünschtes Antwortverhalten spielt für den Unterschied zwischen Objektiv- und Standard-Skala eher keine Rolle, während sozialerwünschtes Antwortverhalten einen Beitrag zur Aufklärung des Abstandes zwischen Objektiv-Kurzskala und Standard-Skala leistet.

In Gruppe A korrelieren die Differenzen zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala auf der Ebene der BIDR-Kurzskala jeweils positiv mit der SD-Skala. Dieser Befund kann bei der Differenz zwischen Objektiv-Kurzskala und Standard-Skala auch für die Subkomponenten der BIDR-Kurzskala bestätigt werden. Dagegen kompensieren sich der Einfluss von SDE- und IM-Komponente auf die Differenz zwischen Objektiv- und Standard-Skala.

Unter der Bedingung, dass die Messintention beider Skalentypen einsichtig war, vergrößert sich der Unterschied zwischen den Objektiv-Skalen und der Standard-Skala infolge sozialerwünschten Antwortverhaltens.

In Gruppe B ist die Differenz zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala jeweils negativ mit der BIDR-Kurzskala korreliert. Dies gilt auch für den Zusammenhang der Differenz zwischen Objektiv-Kurzskala und Standard-Skala mit den Subkomponenten der BIDR-Kurzskala. Dagegen lässt sich bei der Differenz zwischen Objektiv- und Standard-Skala ein Kompensationseffekt von SDE- und IM-Komponente beobachten.

Unter der Bedingung, dass die Messintention der Objektiv-Skalen erfolgreich verschleiert wurde, verringert sich der Unterschied zwischen den Objektiv-Skalen und der Standard-Skala infolge sozialerwünschten Antwortverhaltens.

In Gruppe A sind die Objektiv-Skalen und die Standard-Skala jeweils unterschiedlich stark mit der BIDR-Kurzskala korreliert. Dabei ist die Standard-Skala nicht mit der BIDR-Kurzskala korreliert, während die Objektiv-Skalen jeweils negativ mit der BIDR-Kurzskala korreliert sind. Entsprechend sind die Messwerte der Objektiv-Skalen jeweils stärker durch sozialerwünschtes Antwortverhalten verfälscht als die Messwerte der Standard-Skala. Dabei ist allerdings zu beachten, dass der Befund für die Standard-Skala Folge einer Kompensation von Selbst- und

Fremdtäuschungstendenz ist. Die BIDR-Kurzskala ist jeweils positiv mit den Differenzen zwischen den Messwerten der verschiedenen Skalentypen korreliert.

Die Abstandsvergrößerung zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala scheint vornehmlich auf das Antwortverhalten auf den Objektiv-Skalen zurückzuführen zu sein. So unterschätzen die Probanden ihr wahres Vertrauen auf den Objektiv-Skalen, während sie auf der Standard-Skala aufrichtiger antworten. Letztgenannter Befund kommt allerdings nur durch die wechselseitige Kompensation des Einflusses der Subkomponente sozialer Erwünschtheit auf die Standard-Skala zustande.

In Gruppe B sind die Objektiv-Skalen und die Standard-Skala jeweils unterschiedlich stark mit der BIDR-Kurzskala korreliert. Dabei ist die Objektiv-Kurzskala stärker als die Standard-Skala und die Objektiv-Skala weniger stark als die Standard-Skala durch sozialerwünschtes Antwortverhalten verfälscht. Die Differenzen zwischen den Messwerten der verschiedenen Skalentypen sind jeweils negativ mit der BIDR-Kurzskala korreliert.

Die Abstandsverringerung zwischen den Messwerten der Objektiv-Skala und der Standard-Skala infolge sozialerwünschten Antwortverhaltens scheint vornehmlich auf das Antwortverhalten auf der Standard-Skala zurückzugehen. So unterschätzen die Probanden ihr wahres Vertrauen auf der Standard-Skala etwas stärker, als sie es auf der Objektiv-Skala überschätzen. Die Abstandsverringerung zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala ist dagegen vornehmlich auf das Antwortverhalten auf der Objektiv-Kurzskala zurückzuführen. Hier überschätzen die Probanden ihr wahres Vertrauen deutlich stärker, als sie es auf der Standard-Skala unterschätzen. Die Probanden antworten somit auf der Standard-Skala im Vergleich zur Objektiv-Kurzskala tendenziell aufrichtiger.

In der Zusammenschau bedeutet dies, dass der Unterschied zwischen Objektiv-Kurzskala und Standard-Skala dadurch zustande kommt, dass auf der Standard-Skala tendenziell aufrichtiger geantwortet wird, während die Objektiv-Kurzskala, in Abhängigkeit von der Identifikation der Messintention, das wahre Vertrauen über- bzw. unterschätzt. Der Unterschied zwischen Objektiv- und Standard-Skala ist dagegen bei erfolgreicher Verschleierung der Messintention darauf zurückzuführen, dass auf der Standard-Skala weniger aufrichtig geantwortet wird als auf der Objektiv-Skala. Insofern scheint allenfalls die erfolgreiche Verschleierung der Messintention der Objektiv-Skala einen förderlichen Effekt auf die Messgüte zu haben.

7.2.7.2.2 Differenzierung nach Identifikation der Messintention und Referenzgruppe

Ausgehend vom Radiusproblem der Vertrauensforschung wird die vorangegangene Analyse nochmals anhand der Fälle nachvollzogen, die sowohl für ihre Antworten auf den Objektiv-Skalen als auch auf der Standard-Skala ausschließlich generalisierte Andere bzw. die *outgroup* als Vertrauensreferenz angegeben haben. Auch hier reduziert sich die in die Analyse eingehende Fallzahl aufgrund der zusätzlichen Differenzierungsstufe erheblich.

7. Hauptstudie

Tab. 7.108: Tendenz zu sozialerwünschtem Antwortverhalten (Vertrauensreferenz entspricht *outgroup*)

Gruppe A [Messintention der Objektiv-Skalen erkannt] (N=13)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,6923	4,7692	4,6154
<i>s</i>	,85485	,92681	1,38675
Gruppe B [Messintention der Objektiv-Skalen <u>nicht</u> erkannt] (N=17)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,7647	4,8235	4,6471
<i>s</i>	,90342	1,01460	,93148

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

(Quelle: eigene Berechnungen)

Was die Neigung zu sozialer Erwünschtheit angeht, so ergeben sich zwischen Gruppe A und Gruppe B keine nennenswerten Unterschiede. Dies gilt gleichermaßen für die Subkomponenten der BIDR-Kurzskala.

Tab. 7.109: Bivariate Korrelationen: BIDR_(KS) und PR₂ bzw. PT₂ (N = 13)
[Messintention von PT₂ identifiziert und Vertrauensreferenz entspricht *outgroup* (Gruppe A)]

	PR ₂	PT ₂	PT _{2(KS)}	BIDR _(KS)	SDE	IM
PR ₂	1	---	---	---	---	---
PT ₂	,385	1	---	---	---	---
PT _{2(KS)}	,453	,656*	1	---	---	---
BIDR _(KS)	-,119	-,108	-,532	1	---	---
SDE	-,154	,087	-,504	,639*	1	---
IM	-,235	-,408	-,318	,735**	,120	1

PR₂: Standard-SkalaPT₂: Objektiv-SkalaPT_{2(KS)}: Objektiv-KurzskalaBIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.110: Bivariate Korrelationen: BIDR_(KS) und PR₂ bzw. PT₂ (N = 17)
[Messintention von PT₂ nicht identifiziert und Vertrauensreferenz entspricht *outgroup* (Gruppe B)]

	PR ₂	PT ₂	PT _{2(KS)}	BIDR _(KS)	SDE	IM
PR ₂	1	---	---	---	---	---
PT ₂	,515*	1	---	---	---	---
PT _{2(KS)}	,451	,572*	1	---	---	---
BIDR _(KS)	-,028	-,138	,149	1	---	---
SDE	-,040	-,370	-,041	,566*	1	---
IM	-,087	,134	,063	,786**	,128	1

PR₂: Standard-SkalaPT₂: Objektiv-SkalaPT_{2(KS)}: Objektiv-KurzskalaBIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Aus dem Vergleich von Gruppe A und B geht hervor, dass die Objektiv-Kurzskala in Gruppe B weniger durch sozialerwünschtes Antwortverhalten beeinflusst ist als in Gruppe A. Dieser Befund setzt sich auch für die SDE- und die IM-Komponente fort. Für die Objektiv-Skala scheint sich dagegen zwischen den Gruppen kein nennenswerter Unterschied zu ergeben, was den Einfluss sozialer Erwünschtheit auf das Antwortverhalten angeht.

In Gruppe A verstärken sich die Effekte der Selbst- und Fremdtäuschungskomponente auf der Objektiv-Kurzskala gegenseitig, sodass hier mit erheblich verzerrten Antworten zu rechnen ist. Auf der Objektiv-Skala scheint dagegen vornehmlich die Tendenz zur Fremdtäuschung für Antwortverzerrungen verantwortlich zu sein.

In Gruppe B ist die Objektiv-Skala ähnlich stark durch soziale Erwünschtheit belastet, allerdings ist im Kontrast zu Gruppe A nun vornehmlich die Neigung zur Selbsttäuschung für diesen Effekt verantwortlich. Für die Objektiv-Kurzskala reduziert sich dagegen der Einfluss sozialerwünschten Antwortverhaltens im Vergleich zu Gruppe A erheblich.

Die Messwerte der Objektiv-Kurzskala sind in Gruppe B weniger stark durch sozialerwünschtes Antwortverhalten verfälscht als in Gruppe A. Außerdem ist die Objektiv-Kurzskala in Gruppe B weniger durch soziale Erwünschtheit beeinflusst als in Gruppe A. Die Messwerte der Objektiv-Skala sind in Gruppe A und B ähnlich stark durch sozialerwünschtes Antwortverhalten verfälscht und auch die Beeinflussung der Objektiv-Skala durch soziale Erwünschtheit fällt in beiden Gruppen ähnlich aus. Insofern hat die erfolgreiche Verschleierung der Messintention nur für die Objektiv-Kurzskala eine Reduktion der Verfälschbarkeit durch sozialerwünschtes Antwortverhalten bewirkt.

Die Messwerte der Standard-Skala sind in Gruppe B weniger stark durch soziale Erwünschtheit verzerrt als in Gruppe A. Außerdem ist die Standard-Skala in Gruppe B weniger durch soziale Erwünschtheit beeinflusst als in Gruppe A.

Entsprechend fällt die Anfälligkeit der Standard-Skala für sozialerwünschtes Antwortverhalten in den Gruppen A und B nicht gleich aus bzw. sie scheint unabhängig von der mit der Zieldimension übereinstimmenden Augenscheinvalidität zu sein.

In Gruppe A sind die Messwerte von Objektiv- und Standard-Skala ähnlich stark durch sozialerwünschtes Antwortverhalten verfälscht. Die Messwerte der Objektiv-Kurzskala sind dagegen stärker durch sozialerwünschtes Antwortverhalten verzerrt als die der Standard-Skala. Hierfür scheint insbesondere die wechselseitige Verstärkung von SDE- und IM-Komponente hinsichtlich der Objektiv-Skala verantwortlich zu sein. Die Objektiv-Skalen sind außerdem jeweils stärker durch soziale Erwünschtheit beeinflusst als die Standard-Skala.

Sind die Zieldimensionen bei beiden Skalentypen transparent, werden die Messwerte der Objektiv-Skalen entweder ähnlich stark oder stärker durch sozialerwünschtes Antwortverhalten verfälscht als die der Standard-Skala. Jedoch scheinen beide Skalentypen ähnlich stark durch soziale Erwünschtheit belastet zu sein.

In Gruppe B ist die Standard-Skala weniger stark mit der SD-Skala korreliert als die Objektiv-Skalen. Betrachtet man jedoch die Subkomponenten der BIDR-Kurzskala zeigt sich, dass die Objektiv-Kurzskala und die Standard-Skala ähnlich schwach durch soziale Erwünschtheit beeinflusst sind. Die Objektiv-Skala ist dagegen stärker durch soziale Erwünschtheit beeinflusst als die Standard-Skala und auch ihre Messwerte sind stärker durch sozialerwünschtes Antwortverhalten verzerrt als die der Standard-Skala.

7. Hauptstudie

In Gruppe B sind die Messwerte der Standard-Skala weniger durch sozialerwünschtes Antwortverhalten verzerrt als die der Objektiv-Skalen. Objektiv-Kurzskala und Standard-Skala sind ähnlich schwach durch soziale Erwünschtheit belastet, während die Objektiv-Skala vergleichsweise stark durch soziale Erwünschtheit beeinflusst ist. Die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen macht diese im Vergleich mit der Standard-Skala somit nicht weniger anfällig für sozialerwünschtes Antwortverhalten.

Nachfolgend soll auch hier wieder der Einfluss sozialer Erwünschtheit auf die Mittelwertdifferenz zwischen den verschiedenen Vertrauensskalen betrachtet werden.

Tab. 7.111: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 13)
[Messintention von PT₂ identifiziert und Vertrauensreferenz entspricht *outgroup* (Gruppe A)]

	Diff (PR ₂ - PT ₂)	Diff (PR ₂ - PT _{2(KS)})	BIDR _(KS)	SDE	IM
Diff (PR ₂ - PT ₂)	1	---	---	---	---
Diff (PR ₂ - PT _{2(KS)})	,762**	1	---	---	---
BIDR _(KS)	-,023	,313	1	---	---
SDE	-,220	,254	,639*	1	---
IM	,119	,015	,735**	,120	1

Diff (PR₂ - PT₂): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PR₂ - PT_{2(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.112: Bivariate Korrelationen: BIDR_(KS) und Differenz der Vertrauensskalen (N = 17)
[Messintention von PT₂ nicht identifiziert und Vertrauensreferenz entspricht *outgroup* (Gruppe B)]

	Diff (PR ₂ - PT ₂)	Diff (PR ₂ - PT _{2(KS)})	BIDR _(KS)	SDE	IM
Diff (PR ₂ - PT ₂)	1	---	---	---	---
Diff (PR ₂ - PT _{2(KS)})	,772**	1	---	---	---
BIDR _(KS)	,077	-,130	1	---	---
SDE	,245	-,017	,566*	1	---
IM	-,205	-,138	,786**	,128	1

Diff (PR₂ - PT₂): Differenz zwischen den Summenscores von Objektiv- und Standard-Skala

Diff (PR₂ - PT_{2(KS)}): Differenz zwischen den Summenscores von Objektiv-Kurzskala und Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Aus dem Vergleich der Gruppen A und B geht hervor, dass die Differenz von Objektiv- und Standard-Skala jeweils kaum mit der BIDR-Kurzskala korreliert ist, während die Differenz zwischen Objektiv-Kurzskala und Standard-Skala in beiden Gruppen mit der BIDR-Kurzskala korreliert.

Das bedeutet: Sozialerwünschtes Antwortverhalten leistet kaum einen Beitrag zur Aufklärung des Unterschiedes zwischen Objektiv- und Standard-Skala. Die Differenz zwischen Objektiv-Kurzskala und Standard-Skala scheint dagegen zumindest in Teilen auf den Einfluss sozialer Erwünschtheit zurückzugehen.

In Gruppe A korreliert die Differenz zwischen Objektiv- und Standard-Skala negativ mit der BIDR-Kurzskala, während die Differenz zwischen Objektiv-Kurzskala und Standard-Skala positiv mit der BIDR-Kurzskala korreliert. Dieser Befund lässt sich auch für die Korrelation der Subkomponenten der BIDR-Kurzskala mit der Differenz zwischen Objektiv-Kurzskala und Standard-Skala bestätigen. Dagegen weisen die Korrelationen der Differenz zwischen Objektiv- und Standard-Skala mit den Subkomponenten entgegengesetzte Vorzeichen auf. SDE- und IM-Komponente gleichen sich hier wechselseitig weitestgehend aus.

Unter der Bedingung, dass die Messintention beider Skalentypen offenliegt, verringert sich der Unterschied zwischen der Objektiv- und der Standard-Skala infolge sozialerwünschten Antwortverhaltens, während der Unterschied zwischen Objektiv-Kurzskala und Standard-Skala durch sozialerwünschtes Antwortverhalten vergrößert wird.

In Gruppe B ist die Differenz der Messwerte von Objektiv- und Standard-Skala positiv mit der BIDR-Kurzskala korreliert, während die Differenz zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala negativ mit der BIDR-Kurzskala korreliert ist. Dies gilt auch für die Korrelation der Differenz von Objektiv-Kurzskala und Standard-Skala mit den Subkomponenten der BIDR-Kurzskala. Für die Differenz von Objektiv- und Standard-Skala ergibt sich dagegen eine wechselseitige Kompensation von Selbst- und Fremdtäuschungstendenz.

Unter der Bedingung, dass die Messintention der Objektiv-Skalen erfolgreich verschleiert wurde, vergrößert sich der Unterschied zwischen der Objektiv- und der Standard-Skala infolge sozialerwünschten Antwortverhaltens, während der Unterschied zwischen Objektiv-Kurzskala und Standard-Skala durch sozialerwünschtes Antwortverhalten verringert wird.

In Gruppe A sind die Objektiv-Skalen und die Standard-Skala jeweils unterschiedlich stark mit der BIDR-Kurzskala korreliert, wobei die Korrelationswerte jeweils negative Vorzeichen aufweisen. Dabei ist die Objektiv-Kurzskala stärker durch sozialerwünschtes Antwortverhalten verfälscht als die Standard-Skala, während die Objektiv-Skala etwas weniger durch sozialerwünschtes Antwortverhalten verfälscht wird als die Standard-Skala. Die Korrelationen der BIDR-Kurzskala mit den Differenzen der Messwerte weisen einander entgegengesetzte Vorzeichen auf.

Die Abstandsverringering zwischen den Messwerten der Objektiv- und Standard-Skala scheint vornehmlich auf das Antwortverhalten auf der Standard-Skala zurückzuführen zu sein. Auf beiden Skalen unterschätzen die Probanden ihr wahres Vertrauen, allerdings fällt dieser Effekt auf der Standard-Skala etwas stärker aus als auf der Objektiv-Skala. Für die Abstandsvergrößerung zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala scheint das Antwortverhalten auf der Objektiv-Kurzskala verantwortlich zu sein. Auch hier unterschätzen die Probanden auf beiden Skalen ihr wahres Vertrauen, allerdings wiegt dieser Effekt auf der Objektiv-Kurzskala schwerer als auf der Standard-Skala.

In Gruppe B sind die Objektiv-Skalen und die Standard-Skala jeweils unterschiedlich stark mit der BIDR-Kurzskala korreliert, wobei die Korrelation der Objektiv-Kurzskala mit der BIDR-Kurzskala negativ ausfällt und die verbleibenden Korrelationswerte jeweils ein positives Vorzeichen aufweisen. Die Objektiv-Skalen sind jeweils stärker durch sozialerwünschtes Antwortverhalten verfälscht als die Standard-Skala. Für die Objektiv-Skala lässt sich dabei eine Kompensation der SDE- und der IM-Komponente beobachten. Die Korrelationen der

7. Hauptstudie

Messwertunterschiede mit der BIDR-Kurzskaala weisen jeweils einander entgegengesetzte Vorzeichen auf.

Die Abstandsvergrößerung zwischen Objektiv- und Standard-Skala scheint vornehmlich auf das Antwortverhalten auf der Objektiv-Skala zurückzugehen. So unterschätzen die Probanden ihr wahres Vertrauen zwar auf beiden Skalen, allerdings fällt dieser Effekt für die Objektiv-Skala stärker aus. Die Abstandsverringerung zwischen Objektiv-Kurzskaala und Standard-Skala scheint vornehmlich auf das Antwortverhalten auf der Objektiv-Kurzskaala zurückgeführt werden zu können. Die Probanden überschätzen hier ihr wahres Vertrauen auf der Objektiv-Kurzskaala stärker als sie es auf der Standard-Skala unterschätzen.

In der Zusammenschau bedeutet dies, dass der Unterschied zwischen Objektiv-Kurzskaala und Standard-Skala unabhängig von der Identifikation der Messintention durch das Antwortverhalten auf der Objektiv-Kurzskaala ausgelöst wird. Da der Unterschied zwischen Objektiv-Skala und Standard-Skala bei erfolgreicher Verschleierung der Messintention auf das Antwortverhalten auf der Objektiv-Skala zurückgeht, scheint auch hier die Intransparenz der Zieldimension keinen förderlichen Effekt für die Messgüte zu haben.

7.2.6.3 Gruppe 3: Objektiv

Ebenso wie die Probanden der bisher betrachteten Versuchsgruppen weisen auch die Probanden aus *Versuchsgruppe 3* eine erhöhte Neigung zu sozialerwünschtem Antwortverhalten auf ($\bar{x}_{BIDR} = 4,71$, $s = 0,893$; $\bar{x}_{SDE} = 4,65$, $s = 1,153$; $\bar{x}_{IM} = 4,44$, $s = 1,259$). So zeigt sich im Rahmen der grafischen Analyse, dass die Mehrzahl der Probanden sowohl auf der gesamten SD-Skala als auch für deren Subkomponenten Werte rechts von der Skalenmitte erreichen.

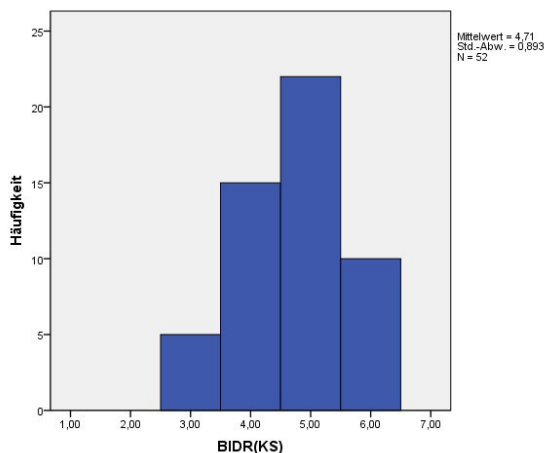


Abb. 7.31: Häufigkeitsverteilung Summenscore BIDR-Kurzskaala („soziale Erwünschtheit“) (mit 1 = geringe Neigung zu SD bis 5 = starke Neigung zu SD) in Versuchsgruppe 3

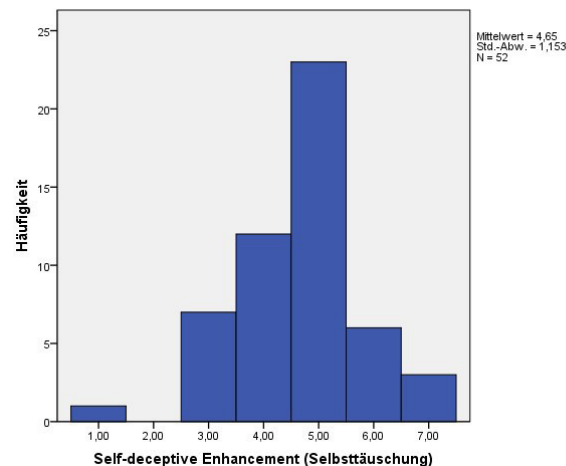


Abb. 7.32: Häufigkeitsverteilung Summenscore SDE-Komponente („Selbsttäuschung“) (mit 1 = geringe Neigung zu SDE bis 5 = starke Neigung zu SDE) in Versuchsgruppe 3

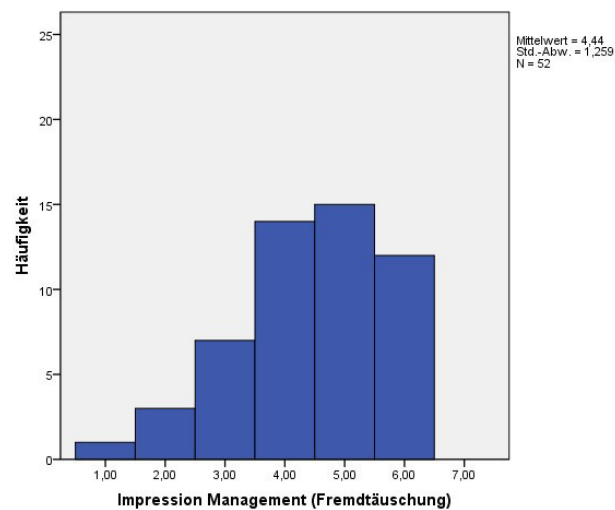


Abb. 7.33: Häufigkeitsverteilung Summenscore IM-Komponente („Fremdtäuschung“) (mit 1 = geringe Neigung zu IM bis 5 = starke Neigung zu IM) in Versuchsgruppe

Tab. 7.113: Bivariate Korrelationen: BIDR_(KS) und PT₃ (N = 52)

	PT ₃	PT _{3(KS)}	BIDR _(KS)	SDE	IM
PT ₃	1	---	---	---	---
PT _{3(KS)}	,760**	1	---	---	---
BIDR _(KS)	,046	-,069	1	---	---
SDE	-,381**	-,393**	,606**	1	---
IM	,423**	,297*	,674**	-,041	1

PT₃: Objektiv-Skala

PT_{3(KS)}: Objektiv-Kurzskala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Während auf der allgemeinen Ebene keine Beeinflussung des Antwortverhaltens auf der Objektiv-Skala bzw. -Kurzskala durch soziale Erwünschtheit zu verzeichnen ist, ergibt sich für die Subkomponenten ein anderes Bild. Die Probanden scheinen dabei sowohl zu Selbsttäuschung als auch zu Fremdtäuschung zu neigen, wobei sich deren Einfluss jeweils diametral entgegensteht. So ist die Neigung zur Selbsttäuschung mit einem Absinken des Generalisierten Vertrauens und die Neigung zur Fremdtäuschung mit einem Anstieg des Generalisierten Vertrauens verbunden. Selbst- und Fremdtäuschung scheinen sich demnach wechselseitig zu kompensieren, was die jeweils geringe Korrelation mit der gesamten BIDR-Kurzskala erklärt. Während der Einfluss der Fremdtäuschung den theoretischen Annahmen entspricht, wonach das Ziel der Probanden darin besteht, nach außen ein möglichst positives Bild abzugeben, d.h. in diesem Fall ein hohes Niveau Generalisierten Vertrauens aufzuweisen, widerspricht die Richtung des Zusammenhangs von Selbsttäuschung und Objektiv-Skalen den theoretischen Annahmen. Diesen zur Folge sollte eine Tendenz dazu, die Realität optimistischer wahrzunehmen als sie objektiv erscheint, einen Anstieg der Vertrauensbereitschaft bewirken. Da es bei der Neigung zur Selbsttäuschung auch um den Schutz des Selbstbildes geht, ist die

7. Hauptstudie

gegensätzliche Steigung von Selbsttäuschung und Generalisiertem Vertrauen vermutlich darauf zurückzuführen, dass die Probanden nicht als naiv oder als zu vertrauensselig erscheinen möchten. Insofern besteht zwar eine Neigung dazu, die Umwelt positiv zu verklären, während allerdings Skepsis gegenüber anderen als gesellschaftlich adäquat gilt. Diese sozialerwünschte Skepsis resultiert schließlich in einem tendenziell geringer ausgeprägten generalisierten Vertrauen.

Die geringe Korrelation von BIDR-Kurzskala und den Objektiv-Skalen täuscht somit über den tatsächlichen Einfluss sozialerwünschten Antwortverhaltens auf die Beantwortung der Objektiv-Skalen in *Versuchsgruppe 3* hinweg. In Wahrheit ist das Antwortverhalten auf den Objektiv-Skalen vergleichsweise stark von sozialer Erwünschtheit betroffen. Aufgrund der wechselseitigen Kompensation der Subkomponenten hat dies aber keinen Einfluss auf die Validität der Vertrauensmessung. Die Probanden haben somit unfreiwillig aufrichtig geantwortet. Insofern bewirkt die Verwendung Objektiver Tests eine Verbesserung der Güte der Vertrauensmessung.

7.2.6.3.1 Differenzierung nach Identifikation der Messintention

Auch im Rahmen von *Versuchsgruppe 3* soll untersucht werden, welche Konsequenzen sich aus der Aufdeckung der Messintention der Objektiv-Skalen für den Effekt sozialerwünschten Antwortverhaltens ergeben.

Tab. 7.114: Tendenz zu sozialerwünschtem Antwortverhalten

Gruppe A [Messintention der Objektiv-Skalen erkannt] (N=11)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,2727	3,9091	4,3636
s	,90453	1,22103	1,28629
Gruppe B [Messintention der Objektiv-Skalen nicht erkannt] (N=39)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,8205	4,8462	4,4359
s	,88472	1,08914	1,29361

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

(Quelle: eigene Berechnungen)

Auch in *Versuchsgruppe 3* neigen die Probanden in beiden der nach Identifikation der Messintention differenzierten Gruppen überdurchschnittlich stark zu sozialerwünschtem Antwortverhalten. Dabei ist diese Tendenz innerhalb der Probandengruppe, die die Messintention der Objektiv-Skalen nicht erkannt hat, etwas stärker ausgeprägt als in der Gruppe, die die Messintention erkannt hat.

Tab. 7.115: Bivariate Korrelationen: BIDR_(KS) und PT₃ (N = 11)
[Messintention von PT₃ identifiziert (Gruppe A)]

	PT ₃	PT _{3(KS)}	BIDR _(KS)	SDE	IM
PT ₃	1	---	---	---	---
PT _{3(KS)}	,174	1	---	---	---
BIDR _(KS)	,202	,124	1	---	---
SDE	-,219	,037	,477	1	---
IM	,590	,245	,680*	-,232	1

PT₃: Objektiv-Skala

PT_{3(KS)}: Objektiv-Kurzskala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.116: Bivariate Korrelationen: BIDR_(KS) und PT₃ (N = 39)
[Messintention von PT₃ nicht identifiziert (Gruppe B)]

	PT ₃	PT _{3(KS)}	BIDR _(KS)	SDE	IM
PT ₃	1	---	---	---	---
PT _{3(KS)}	,882**	1	---	---	---
BIDR _(KS)	,081	-,075	1	---	---
SDE	-,372*	-,474**	,599**	1	---
IM	,399*	,319*	,691**	-,007	1

PT₃: Objektiv-Skala

PT_{3(KS)}: Objektiv-Kurzskala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Der Vergleich von Gruppe A und Gruppe B zeigt auf der Ebene der BIDR-Kurzskala, dass die Objektiv-Skalen in Gruppe B weniger durch sozialerwünschtes Antwortverhalten beeinflusst sind. Betrachtet man allerdings die Subkomponenten der BIDR-Kurzskala scheint die Objektiv-Skala in beiden Gruppen ähnlich stark durch soziale Erwünschtheit beeinflusst zu sein, während sich der Effekt für die Objektiv-Kurzskala in Gruppe B sogar verstärkt.

In Gruppe A ist die Objektiv-Skala stärker von sozialerwünschtem Antwortverhalten betroffen als die Objektiv-Kurzskala. Die SDE- und die IM-Komponente kompensieren sich allerdings teilweise wechselseitig. Die Objektiv-Kurzskala ist im Vergleich etwas weniger durch soziale Erwünschtheit beeinflusst, da sie kaum durch Selbsttäuschung beeinflusst wird.

In Gruppe B zeigt sich zwar für beide Objektiv-Skalen auf der Ebene der BIDR-Kurzskala eine geringe Belastung durch sozialerwünschtes Antwortverhalten, allerdings lässt sich dieser Befund mit Blick auf die SDE- und die IM-Komponente nicht bestätigen. So kompensieren sich die Selbst- und Fremdtäuschungstendenz bei beiden Skalen wechselseitig. Dies hat zur Folge, dass die Probanden insgesamt zwar unbewusst aufrichtiger geantwortet haben, dies aber nur aufgrund der Gegenläufigkeit der Antworttendenzen geschieht.

Es lässt sich daher schließen, dass die Objektiv-Skalen in Gruppe B nicht weniger stark mit der SD-Skala korreliert sind als in Gruppe A. Ausgehend von den Subkomponenten ergibt sich hinsichtlich der Objektiv-Skala eher kein Unterschied zwischen den Gruppen, während die Objektiv-Kurzskala in Gruppe B eher stärker durch sozialerwünschtes Antwortverhalten beeinflusst ist. Die erfolgreiche Verschleierung der Messintention bewirkt allerdings, dass die Messwerte insgesamt weniger durch sozialerwünschtes Antwortverhalten verfälscht sind.

7.2.7.3.2 Differenzierung nach Identifikation der Messintention und Referenzgruppe

Auch in *Versuchsgruppe 3* soll dem Radiusproblem Rechnung getragen werden. Entsprechend gehen in die nachfolgende Analyse ausschließlich solche Fälle ein, die generalisierte Andere bzw. die *outgroup* als Vertrauensreferenz angegeben haben. Hierbei ist zu beachten, dass sich die Fallanzahl aufgrund der Auswahlparameter teilweise drastisch reduziert.

7. Hauptstudie

Tab. 7.117: Tendenz zu sozialerwünschtem Antwortverhalten

<i>Gruppe A [Messintention der Objektiv-Skalen erkannt] (N=4)</i>			
	BIDR_(KS)	SDE	IM
\bar{x}	4,2500	3,5000	4,5000
<i>s</i>	,95743	1,73205	1,73205
<i>Gruppe B [Messintention der Objektiv-Skalen nicht erkannt] (N=27)</i>			
	BIDR_(KS)	SDE	IM
\bar{x}	4,8148	4,9630	4,2222
<i>s</i>	,83376	,93978	1,33973

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

(Quelle: eigene Berechnungen)

Sowohl die Probanden, die die Messintention der Objektiv-Skalen identifiziert haben, als auch die Probanden, die dies nicht vermochten, weisen eine überdurchschnittliche Tendenz zu sozialerwünschtem Antwortverhalten auf.

Tab. 7.118: Bivariate Korrelationen: BIDR_(KS) und PT₃ (N = 4)

[Messintention von PT₃ identifiziert und Vertrauensreferenz entspricht outgroup (Gruppe A)]

	PT₃	PT_{3(KS)}	BIDR_(KS)	SDE	IM
PT₃	1	---	---	---	---
PT_{3(KS)}	,522	1	---	---	---
BIDR_(KS)	,636	,522	1	---	---
SDE	-,302	,577	,302	1	---
IM	,905	,192	,704	-,444	1

PT₃: Objektiv-SkalaPT_{3(KS)}: Objektiv-KurzskalaBIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Tab. 7.119: Bivariate Korrelationen: BIDR_(KS) und PT₃ (N = 27)

[Messintention von PT₃ nicht identifiziert und Vertrauensreferenz entspricht outgroup (Gruppe B)]

	PT₃	PT_{3(KS)}	BIDR_(KS)	SDE	IM
PT₃	1	---	---	---	---
PT_{3(KS)}	,916**	1	---	---	---
BIDR_(KS)	,129	,013	1	---	---
SDE	-,380	-,502**	,531**	1	---
IM	,452*	,344	,727**	-,054	1

PT₃: Objektiv-SkalaPT_{3(KS)}: Objektiv-KurzskalaBIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*SDE: *self-deceptive enhancement* (Selbsttäuschung)IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Der Vergleich der Gruppen A und B zeigt, dass die Objektiv-Skalen in Gruppe A augenscheinlich stärker durch sozialerwünschtes Antwortverhalten beeinflusst sind, als in Gruppe B. Betrachtet man allerdings die Subkomponenten der BIDR-Kurzskala, so lässt sich kaum ein Unterschied feststellen.

7. Hauptstudie

In Gruppe A kompensieren sich der Einfluss von SDE- und IM-Komponente auf die Objektiv-Skala teilweise wechselseitig. Für die Objektiv-Kurzska lässt sich dagegen ein Verstärkungseffekt der Subkomponenten feststellen. Insgesamt ist die Tendenz zu sozialerwünschtem Antwortverhalten stark ausgeprägt.

In Gruppe B lässt sich für beide Objektiv-Skalen ein Kompensationseffekt der Subkomponenten der BIDR-Kurzska feststellen. Die Probanden antworten dadurch in Gruppe B insgesamt unbewusst aufrichtiger. Die Belastung durch sozialerwünschtes Antwortverhalten bewegt sich allerdings auf einem ähnlichen Niveau wie in Gruppe A.

Es lässt sich daher schließen, dass die Objektiv-Skalen in Gruppe B nicht weniger stark mit der SD-Skala korreliert sind als in Gruppe A. Vielmehr scheint sich hinsichtlich des Einflusses sozialerwünschten Antwortverhaltens kein Unterschied zu ergeben. Allerdings bewirkt die erfolgreiche Verschleierung der Messintention, dass die Messwerte weniger stark durch sozialerwünschte Antworten verfälscht werden.

7.2.7.4 Gruppe 4: Standard

Ebenso wie die Probanden in den vorangegangenen Versuchsgruppen neigen auch die Probanden in *Versuchsgruppe 4* stark zu sozialerwünschtem Antwortverhalten ($\bar{x}_{BIDR} = 4,88$, $s = 0,758$; $\bar{x}_{SDE} = 4,96$, $s = 1,028$; $\bar{x}_{IM} = 4,65$, $s = 1,297$). Dies wird umso deutlicher, wenn man die zugehörigen Histogramme betrachtet. So erreichen die Probanden mehrheitlich Werte rechts von der Mittelkategorie der SD-Skala. Gleiches gilt für deren Subkomponenten. Dabei erscheint insbesondere die Neigung zur Selbsttäuschung besonders stark ausgeprägt, was in einer starken Rechtsverschiebung der BIDR-Kurzska resultiert.

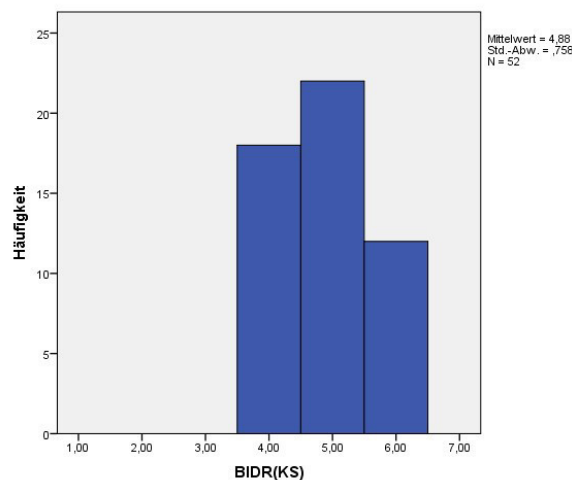


Abb. 7.34: Häufigkeitsverteilung Summenscore BIDR-Kurzska („soziale Erwünschtheit“) (mit 1 = geringe Neigung zu SD bis 5 = starke Neigung zu SD) in Versuchsgruppe 4

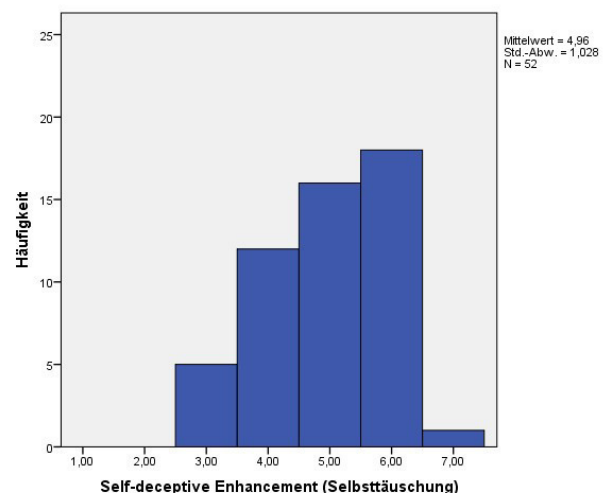


Abb. 7.35: Häufigkeitsverteilung Summenscore SDE-Komponente („Selbsttäuschung“) (mit 1 = geringe Neigung zu SDE bis 5 = starke Neigung zu SDE) in Versuchsgruppe 4

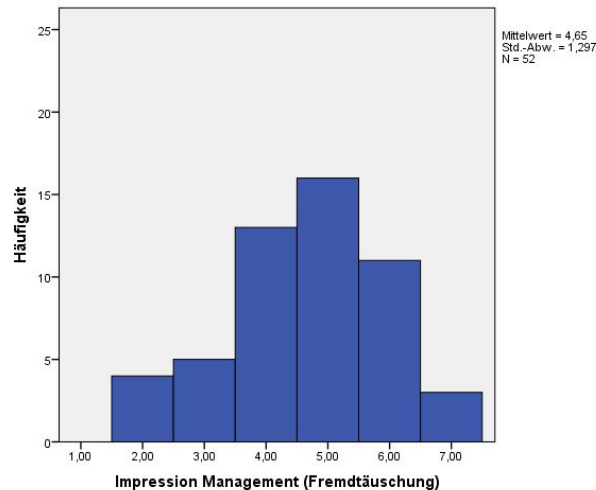


Abb. 7.36: Häufigkeitsverteilung Summenscore IM-Komponente („Fremdtäuschung“) (mit 1 = geringe Neigung zu IM bis 5 = starke Neigung zu IM) in Versuchsgruppe 4

Tab. 7.120: Bivariate Korrelationen: BIDR_(KS) und PT₄ (N = 52)

	PT ₄	BIDR _(KS)	SDE	IM
PT ₄	1	---	---	---
BIDR _(KS)	-,027	1	---	---
SDE	-,108	,548**	1	---
IM	,083	,677**	-,054	1

PT₄: Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Auf der Ebene der BIDR-Skala lässt sich keine Beeinflussung des Antwortverhaltens auf der Standard-Skala durch soziale Erwünschtheit beobachten. Betrachtet man allerdings die Subkomponenten, so scheinen sich auch in *Versuchsgruppe 4* *self-deceptive enhancement* und *impression management* wechselseitig weitestgehend zu kompensieren. Während die IM-Komponente in Richtung der theoretischen Annahmen wirkt, weist die SDE-Komponente in die umgekehrte Richtung. Diese Konstellation lässt sich dabei aller Voraussicht nach in gleicher Art und Weise erklären wie bereits schon in *Versuchsgruppe 3*. So scheinen innerhalb der Probanden zwei widerstreitende Tendenzen wirksam zu sein. Fasst man Selbsttäuschung als Versuch auf, sein Selbstbild zu bewahren, und bewertet gleichzeitig ein ausgeprägtes Generalisiertes Vertrauen bzw. den arglosen Umgang mit anderen als Naivität, wird deutlich, warum die SDE-Komponente negativ mit Generalisiertem Vertrauen verbunden ist. So gilt Naivität gemeinhin als ein Ausdruck von Einfältigkeit, was eine gesamtgesellschaftlich als negativ besetzte Eigenschaft darstellt. Die Probanden hatten somit ein Interesse daran, ihr Selbstbild davor zu schützen, als naiv und vertrauensselig bzw. unkritisch gegenüber anderen zu gelten. Entsprechend ergeben sich mit steigender Tendenz zur Selbsttäuschung niedrigere Vertrauenswerte.

Fremdtäuschung scheint derweil eine untergeordnete bzw. vernachlässigbare Rolle für das Antwortverhalten auf der Standard-Skala zu spielen. Auch wenn sich der Betrag der

7. Hauptstudie

Korrelationswerte von SDE- und IM-Komponente und Standard-Skala kaum unterscheidet und beide ähnlich niedrig ausfallen, scheint die Veranlagung zur Selbsttäuschung von größerer Bedeutung für das Antwortverhalten auf der Standard-Skala zu sein, als die Neigung nach außen ein möglichst günstiges Bild seines Selbst abzugeben.

Ausgehend von diesen Ergebnissen kann die Vertrauensmessung innerhalb von Versuchsgruppe 4 als weitestgehend verzerrungsfrei interpretiert werden. Mit anderen Worten: Die Standard-Skala ermöglicht hier eine valide Abbildung Generalisierten Vertrauens.

7.2.7.4.1 Differenzierung nach Referenzgruppe

Nachfolgend wird innerhalb von *Versuchsgruppe 4* geprüft, inwiefern sich die Korrelation der Standard-Skala mit der BIDR-Kurzskala verhält, wenn man das Radiusproblem berücksichtigt. Da die Messintention der Standard-Skala grundsätzlich für die Probanden einsehbar war, erfolgt keine Differenzierung entlang der Identifikation der Zieldimension.

Tab. 7.121: Tendenz zu sozialerwünschtem Antwortverhalten

Referenzgruppe entspricht outgroup (N=32)			
	BIDR _(KS)	SDE	IM
\bar{x}	4,9063	5,0938	4,5625
s	,77707	,96250	1,31830

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

(Quelle: eigene Berechnungen)

Die in *Versuchsgruppe 4* zusammengefassten Probanden, die generalisierte Andere als Vertrauensreferenz angegeben haben, neigen überdurchschnittlich zu sozialerwünschtem Antwortverhalten.

Tab. 7.122: Bivariate Korrelationen: BIDR_(KS) und PT₄ (N = 32)

[Referenzgruppe entspricht outgroup]

	PT ₄	BIDR _(KS)	SDE	IM
PT ₄	1	---	---	---
BIDR _(KS)	,086	1	---	---
SDE	-,189	,443*	1	---
IM	,247	,714**	-,119	1

PT₄: Standard-Skala

BIDR_(KS): gekürzte Version des *Balanced Inventory of Desirable Responding*

SDE: *self-deceptive enhancement* (Selbsttäuschung)

IM: *impression management* (Fremdtäuschung)

*p<.05 (zweiseitig), **p<.01 (zweiseitig)

(Quelle: eigene Berechnungen)

Auf der Ebene der BIDR-Kurzskala ist die Beeinträchtigung des Antwortverhaltens auf der Standard-Skala durch soziale Erwünschtheit eher gering. Betrachtet man allerdings die Ebene der Subkomponenten, zeigt sich, dass sich die Selbst- und Fremdtäuschungstendenz wechselseitig weitestgehend ausgleichen. Die Probanden scheinen somit eher unbewusst aufrichtiger geantwortet zu haben. Tatsächlich kommt dieses Antwortverhalten aber durch die gegensätzliche Polung der Subkomponenten zustande.

Die Messwerte der Standard-Skala sind somit insgesamt weniger durch sozialerwünschte Antworten verfälscht, was erst durch den Einfluss sozialerwünschten Antwortverhaltens erreicht wird.

7.2.7.5 Wechselseitige Kompensation von SDE- und IM-Komponente

Ein besonders interessanter Befund der Analyse des Einflusses sozialer Erwünschtheit auf das Antwortverhalten auf den verschiedenen Vertrauensskalen ist der eines Kompensationseffektes der Selbst- und Fremdtäuschungskomponente. Dieser ist derart gestaltet, dass sich SDE- und IM-Komponente in manchen Fällen wechselseitig diametral entgegenwirken. Die Folge ist, dass zwar sowohl die Objektiv-Skalen als auch die Standard-Skala teilweise hochgradig durch sozialerwünschtes Antwortverhalten beeinflusst sind, durch die Kompensation der Wirkrichtungen beider SD-Komponenten insgesamt aber tendenziell aufrichtig geantwortet wurde. Mit anderen Worten: Die Verzerrungen infolge von Selbst- und Fremdtäuschung gleichen sich gegenseitig aus, sodass in Summe der wahre Wert bzw. ein annähernd dem wahren Wert entsprechender Messwert resultiert.

Über alle Gruppen und alle Differenzierungsstufen hinweg betrachtet tritt dieser Effekt vornehmlich bei den Objektiv-Skalen auf. So lässt sich für die Objektiv-Skala in 11 von 15 Fällen ein Kompensationseffekt beobachten, wobei sich in 4 Fällen SDE- und IM-Komponente nahezu vollständig ausgleichen und in 6 Fällen zumindest eine Reduktion der Verzerrung des wahren Wertes beobachtet werden kann. Für die Objektiv-Kurzskala ergibt sich in 10 von 15 Fällen ein Kompensationseffekt. Dabei gleichen sich die Selbst- und Fremdtäuschungskomponente in 6 Fällen nahezu vollständig aus und in 4 Fällen ist zumindest eine geringere Verzerrung des wahren Wertes zu verzeichnen. Bei der Standard-Skala tritt nur in 6 von 13 Fällen⁸⁹ ein Kompensationseffekt auf. Dabei kompensieren sich die Effekte von Selbst- und Fremdtäuschung in 5 Fällen nahezu komplett, während in einem Fall zumindest eine Reduktion der Verzerrungen infolge sozialer Erwünschtheit festgestellt werden kann.

Insofern scheint das Design der Objektiv-Skalen die Probanden dazu zu verleiten, unbewusst aufrichtiger zu antworten bzw. das Design der Objektiv-Skalen verhindert zielgerichtete Antwortverzerrungen.

Differenziert man nach Identifikation der Messintention der Objektiv-Skalen und betrachtet dabei ausschließlich die Fälle, die generalisierte Andere als Vertrauensreferenz angeben, zeigt sich, dass der Kompensationseffekt bei der Objektiv-Kurzskala häufiger auftritt, wenn die Verschleierung der Zieldimension erfolgreich war.

Es wird daher vermutet, dass es den Probanden tendenziell weniger möglich ist, zielgerichtete Verfälschungen ihres Antwortverhaltens herbeizuführen, wenn die „demand characteristics“ der Messung nicht klar definiert sind. Mit anderen Worten: Die Anwendung insbesondere der Objektiv-Kurzskala kann zwar nicht verhindern, dass die Probanden sozialerwünscht antworten, allerdings scheinen bewusste Verzerrungen infolge sozialerwünschten Antwortverhaltens reduziert werden zu können. Durch die geringe Augenscheinvalidität der Items der Objektiv-Kurzskala wird dabei eine Definition der Zieldimension verhindert, was ein systematisches sozialerwünschtes Antwortverhalten erschwert.

⁸⁹ Da die Messintention der Standard-Skala prinzipiell transparent gestaltet ist und in *Versuchsgruppe 4* ausschließlich die Standard-Skala präsentiert wurde, konnte für diese Versuchsgruppe keine Differenzierung nach Identifikation der Messintention realisiert werden.

8. Fazit

Ziel dieses Forschungsprojektes war es zu untersuchen, ob und inwiefern die Integration direkter und indirekter Messverfahren im Rahmen von Befragungen eine Reduktion reaktiver Effekte gegenüber dem klassischen Fragebogendesign herbeiführen kann. Die leitende Idee hierbei war, dass eine Kombination dieser beiden Verfahrensgruppen eine wechselseitige Kompensation von deren Vor- und Nachteilen ermöglicht.

Als herausragende positive Eigenschaften von direkten Messverfahren wurden dabei die Möglichkeit der Adaption auf unterschiedlichste Inhalte, die Option der spezifischen Abfrage einer Zieldimension, die vergleichsweise schnelle Konzeptionierung entsprechender Untersuchungsinstrumente und die einfache Umsetzbarkeit dazugehöriger Studiendesigns ausgemacht. Besonders hervorstechende und vornehmlich aus der Transparenz der jeweiligen Messintention resultierende negative Eigenschaft von direkten Messverfahren ist ihre Anfälligkeit für sogenannte reaktive Effekte, aus denen sich Messwertverzerrungen ergeben können, die wiederum die Validität der Ergebnisse bedrohen.

Indirekte Messverfahren zeichnen sich vor allem durch eine geringe Anfälligkeit für reaktive Effekte aus. Die Messwerte, die mit Hilfe solcher Erhebungsmethoden gewonnen werden, sind dementsprechend weniger stark verfälscht. Daneben sind indirekte Messverfahren dazu in der Lage, Facetten eines Forschungsgegenstandes abzubilden, die mittels direkter Messverfahren nicht erfasst werden können. Großer Nachteil indirekter Messverfahren ist jedoch, dass ihre Messkonzepte und -prinzipien oft nur auf bestimmte theoretische Konstrukte oder empirische Phänomene angewendet werden können. In vielen Fällen ist schon eine Operationalisierung, manchmal aber auch nur die praktische Umsetzung einer Messvorschrift mit erheblichem Aufwand verbunden.

Um einen Synergieeffekt zu erreichen, galt es somit die Flexibilität der direkten Messverfahren mit der weitestgehenden Unverfälschbarkeit der indirekten Messverfahren zu verbinden. Als Ansatzpunkt wurden dabei die von CATTELL entwickelten Objektiven Tests und im Speziellen deren Fragebogenvariante gewählt. Daraus ergaben sich für eine Messmethode, die die Vorzüge direkter und indirekter Messverfahren vereint und dabei gleichzeitig ihre Nachteile ausgleicht, die folgenden Designeckpunkte: Erfassung von Verhalten und Eigenschaften von Individuen in (1) einer normierten Situation bzw. unter standardisierten Bedingungen, (2) unter Vermeidung bzw. ohne Notwendigkeit einer Selbstbeurteilung der Individuen und (3) ohne dass die verwendete Messvorschrift einen Hinweis auf die Messintention des Untersuchungsinstrumentes gibt bzw. die eingesetzten Items eine mit der Zieldimension übereinstimmende Augenscheinvalidität aufweisen.

8.1 Reliabilität und Homogenität

Um zu prüfen, ob das Resultat einer Kombination direkter und indirekter Messverfahren, welches nachfolgend als Hybridklasse bezeichnet werden soll, tatsächlich dazu in der Lage ist, Messwertverfälschungen infolge reaktiver Effekte zu minimieren, musste zunächst ein entsprechendes Untersuchungsinstrument entwickelt werden.

Als inhaltliches Vehikel bzw. Proxy für Verhaltensformen oder individuelle Eigenschaft im Allgemeinen wurde das Merkmal „Generalisiertes Vertrauen“ gewählt. Vorteil dieses theoretischen Konstrukts war es, dass mit der *Trust in People-Scale* ein Standardinstrumentarium existierte, das in allen großen Bevölkerungsbefragungen (z.B.

ALLBUS, SOEP, ESS und GSS) in gleicher Art und Weise oder mit leichten Abwandlungen zur Erfassung von Generalisiertem Vertrauen eingesetzt wird und wurde. Die Selektion Generalisierten Vertrauens als Zieldimension der Hybridklasse schafft somit die Voraussetzung für eine Kontrastierung der Hybridklasse mit einem allgemein anerkannten Messinstrument im klassischen Fragebogendesign. Außerdem zeigen empirische Untersuchungen, dass die Abbildung Generalisierten Vertrauens mit Hilfe der *Trust in People-Scale* anfällig für reaktive Effekte und im Speziellen sozialerwünschtes Antwortverhalten ist. Anhand des Merkmals „Generalisiertes Vertrauen“ sollte sich demnach gut detektieren lassen, ob die Verwendung der Hybridklasse eine Verbesserung der Messgüte im Vergleich mit dem Standardinstrument bewirkt.

An dieser Stelle sei allerdings angemerkt, dass prinzipiell auch jedes andere theoretische Konstrukt als Zieldimension infrage kommen würde. Die Anwendung der Hybridklasse ist insofern nicht exklusiv auf das Merkmal „Generalisiertes Vertrauen“ beschränkt. Generalisiertes Vertrauen stellt vielmehr eine Art Platzhalter für die Vielzahl sozialer Variablen dar.

Die Entwicklung eines Messinstruments zur Erfassung Generalisierten Vertrauens, das den Spezifika der Hybridklasse entspricht, erfolgte im Rahmen einer, in diesem Forschungsprojekt als Pilotstudie bezeichneten, Querschnittsuntersuchung. Hauptbestandteil der Hybridklasse sind subtile Items, die eine geringe Augenscheinvalidität bezüglich der tatsächlichen Messintention aufweisen. Voraussetzung für die Formulierung solcher Items ist eine detaillierte Kenntnis der Definition bzw. theoretischen Verfasstheit sowie der Indikatoren und Korrelate des jeweils interessierenden Merkmals oder theoretischen Konstrukts. Nur wenn man diese Parameter kennt, lassen sich Items konstruieren, die einen indirekten Zugang zur Zieldimension ermöglichen. Entsprechend erfolgte zunächst eine umfassende theoretische Aufarbeitung der Charakteristika Generalisierten Vertrauens, deren Ergebnisse die Grundlage für die Operationalisierung darstellten. Im Ergebnis wurden Items formuliert, die augenscheinlich allenfalls einen mittelbaren Bezug zur Zieldimension Generalisiertes Vertrauen aufwiesen. Dies geschah im Wesentlichen unter Rückgriff auf Korrelate Generalisierten Vertrauens und mit Vertrauen verwandte Aspekte oder assoziiertes Verhalten sowie der Vermeidung von Begriffen, die das Wort „Vertrauen“ oder entsprechende Wortstämme enthielten. Auch wurde nach Möglichkeit auf den Gebrauch von zugehörigen Synonymen verzichtet, um den Probanden keinen Hinweis auf die Messintention zu geben. Zusätzlich wurde den Items eine Coverstory vorangestellt, die von der tatsächlichen Messintention ablenken bzw. zu deren erfolgreicher Verschleierung beitragen sollte.

Resultat dieser Bemühungen war ein 42 Items umfassende Rohskala, die insgesamt 452 Probanden zur Bearbeitung vorgelegt wurde. Aufgrund von fehlenden Werten konnten für die weiteren Analysen allerdings nur 220 Fälle berücksichtigt werden.

Im Zuge einer Item- und Reliabilitätsanalyse wurde nun die Item-Anzahl der Rohskala reduziert. Im Rahmen der Itemanalyse wurden nur solche Items beibehalten, die möglichst hohe Item-Trennschärfen und/oder mittlere Item-Schwierigkeiten aufwiesen. Das bedeutet: Es wurden nur solche Items in der Skala belassen, die die Zieldimension „Generalisiertes Vertrauen“ besonders gut abbilden konnten und/oder gut zwischen verschiedenen starken Ausprägungen Generalisierten Vertrauens differenzieren konnten. Dabei stellte die Trennschärfe das primäre und die Schwierigkeit das sekundäre Selektionskriterium dar. Das heißt, es wurden Items mit niedrigen Trennschärfen ausgeschlossen, auch wenn sie mittlere bzw. günstige Schwierigkeiten aufwiesen, und es wurden Items mit vergleichsweise niedrigen oder vergleichsweise hohen Item-Schwierigkeiten beibehalten, wenn sie ausreichend hohe

8. Fazit

Trennschärfen besaßen, um auch zwischen extremen Merkmalsausprägungen trennen zu können. Auf diese Weise sollte eine möglichst homogene Skala entstehen, die nur eine einzelne Dimension („Generalisiertes Vertrauen“) abbildet.

Aus der Itemanalyse resultierten zwei Skalen, die nachfolgend als Objektiv-Skala und Objektiv-Kurzskala bzw. Objektiv-Skalen bezeichnet werden. Die Objektiv-Skala umfasste insgesamt 17 Items, die mindestens Item-Trennschärfen oberhalb des Schwellenwertes für niedrige Trennschärfen ($r_{i(x-i)} \geq .30$) aufwiesen, wobei auch extreme Item-Schwierigkeiten erlaubt waren. In die Objektiv-Kurzskala wurden dagegen ausschließlich die Items mit den höchsten Trennschärfen bzw. „guten“ Trennschärfen ($r_{i(x-i)} = .40-.70$) und mittleren Item-Schwierigkeiten aufgenommen, sodass die Objektiv-Kurzskala nur noch 7 der ursprünglich 17 Items der Objektiv-Skala enthielt.

Im Rahmen der Reliabilitätsanalyse wurden die Messgenauigkeit der Objektiv-Skalen bestimmt und gegebenenfalls weitere Items aus den Skalen eliminiert, wenn dies sich förderlich auf Ausprägung der Reliabilität auswirkte. Der Umfang der Objektiv-Skala wurde im Zuge der Alpha-Maximierung auf 14 Items reduziert, wobei die Skala ein Cronbachs Alpha von $\alpha = .801$ erreichte. Die per Odd-Even-Methode bestimmten Split-Half-Koeffizienten lagen mit $r_{tt}^{90} = .799$ und $\rho_{tt}^{91} = .797$ leicht unterhalb des Wertes für die Innere Konsistenz, während die per Methode der Item-Zwillinge bestimmten Split-Half-Koeffizienten mit $r_{tt} = .818$ und $\rho_{tt} = .818$ leicht oberhalb von Cronbachs Alpha rangierten. Insgesamt ist die Reliabilität der Objektiv-Skala somit mit „gut“ zu bewerten.

Aus der Objektiv-Kurzskala wurden im Laufe der Alpha-Maximierung keine weiteren Items entfernt. Die Skala erreicht dabei ein Cronbachs Alpha von $\alpha = .728$. Die per Odd-Even-Methode bestimmten Split-Half-Koeffizienten erreichten Wert von $r_{tt} = .721$ und $\rho_{tt} = .705$, während die per Methode der Item-Zwillinge bestimmten Split-Half-Koeffizienten Werte von $r_{tt} = .743$ und $\rho_{tt} = .726$ annahmen. Es ergaben sich somit ähnliche Resultate wie für die Innerer Konsistenz der Skala. Insgesamt ist die Reliabilität der Objektiv-Kurzskala mit „akzeptabel“ zu bewerten.

Der Skalenvergleich zeigt, dass die Reliabilität der Objektiv-Skala höher ausfällt als die der Objektiv-Kurzskala. Dieser Umstand weist nicht unbedingt auf eine geringere Messgenauigkeit der Objektiv-Kurzskala hin, sondern rührt vermutlich von den Unterschieden in der Skalenlänge her. So wird eine Messung mit Hilfe einer größeren Anzahl von Items oft automatisch präziser, da mehr Facetten einer Zieldimension abgefragt werden. Die Reliabilitäten bzw. die Messgenauigkeit der Objektiv-Skalen ist vergleichbar mit alternativen Vertrauensskalen. Dort werden Reliabilitätswerte zwischen .57 und .92 (sowohl Innere Konsistenz als auch Split-Half-Reliabilität) erreicht. Zu bemerken ist allerdings, dass die Messgenauigkeit der Objektiv-Skalen jeweils unterhalb der Messgenauigkeit der *Faith in People-Scale* ($\alpha = .92$) liegt. Diese Skala kann als vergleichbar mit dem Standardinstrument der Vertrauensmessung (*Trust in People-Scale*), für das allenfalls eine rudimentäre Spezifizierung der Inneren Konsistenz durch die Bestimmung der Item-Interkorrelationen ($r = .48-.54$) vorliegt, angesehen werden. So enthält die *Faith in People-Scale* im Wesentlichen die gleichen Items wie die *Trust in People-Scale*, wenn auch teilweise in abgewandelter Form, ist allerdings im Vergleich um zwei zusätzliche Items erweitert. Die Item-Interkorrelationen der Objektiv-Skala nehmen Werte von $r = .047-.416$ an, während die Item-Interkorrelationen der Objektiv-Kurzskala Werte von $r = .152-.416$ annehmen. Das bedeutet: Die Items der Objektiv-Skalen weisen einen geringeren Zusammenhang auf als

⁹⁰ Spearman-Brown-Koeffizient

⁹¹ Guttman-Split-Half-Koeffizient

die Items der *Trust in People-Scale*. Die Objektiv-Skalen scheinen demnach eine differenziertere Abbildung der Zieldimension „Generalisiertes Vertrauens“ zu leisten, die im Vergleich zur *Trust in People-Scale* aber auf Kosten der Messpräzision gehen könnte. Dabei scheint allerdings die kompaktere Objektiv-Kurzskaala im Vergleich zur Objektiv-Skala homogener zu sein, da die einzelnen Items dort im Durchschnitt einen stärkeren Zusammenhang aufweisen. Entsprechend ist anzunehmen, dass die Objektiv-Kurzskaala eine im Kontrast zur Objektiv-Skala präzisere Abbildung Generalisierten Vertrauens liefert.

Im Rahmen einer Hauptachsenanalyse wurde außerdem eine rudimentäre Prüfung auf Eindimensionalität der aus Item- und Reliabilitätsanalyse resultierenden Objektiv-Skalen durchgeführt. Für die Objektiv-Skala ergab sich dabei in erster Instanz eine mehrdimensionale Faktorstruktur mit den Dimensionen *Vertrauensbereitschaft*, *Misanthropie*, *Fairness* und *Hilfsbereitschaft*. Die Ermittlung der Faktoren zweiter Ordnung ergab dagegen nur noch eine einzelne Dimension, die als Generalisiertes Vertrauen bezeichnet wurde. Auch für die Objektiv-Kurzskaala ergab sich bei der Extraktion der Faktoren erster Ordnung eine multifaktorielle Lösung mit den Dimensionen *Vertrauensbereitschaft* und *Fairness*. Eine Bestimmung der Faktoren höherer Ordnung war an dieser Stelle aufgrund der geringen Anzahl von Variablen, die in eine solche Analyse Eingang finden würden, nicht möglich. Allerdings ergab sich zwischen den extrahierten Faktoren ein starker positiver Zusammenhang ($r = .621$), sodass ein übergeordneter gemeinsamer Faktor als höchstwahrscheinlich angenommen werden kann.

Die einzelnen Items der *Trust in People-Scale* sollen die Dimensionen *Vertrauensbereitschaft*, *Fairness* und *Hilfsbereitschaft* abbilden, während die Skala insgesamt den individuellen Vertrauenslevel der Probanden anzeigt. Aus der explorativen Faktorenanalyse resultieren für die Objektiv-Skalen zunächst jeweils Faktorenstrukturen, die in weiten Teilen der Faktorstruktur der *Trust in People-Scale* entsprechen. Auf inhaltlicher Ebene legt dies nahe, dass es sich bei *Vertrauensbereitschaft*, *Fairness* und *Hilfsbereitschaft* um die wesentlichen Kernaspekte Generalisierten Vertrauens handelt. Zusätzlich ergibt sich hieraus ein Hinweis auf die Validität der Messergebnisse der Objektiv-Skalen. So lassen sich aus der Parallelität der Faktorstrukturen Rückschlüsse auf die Zieldimension der Skalen ableiten. Da für die *Trust in People-Scale* aufgrund vorangegangener empirischer Studien und theoretischer Diskussionen feststeht, dass sie Generalisiertes Vertrauen abbildet, scheint sich aus der weitestgehenden Äquivalenz der Faktorstrukturen der Skalen, für die aus der Theorie abgeleitete Annahme, dass auch die Objektiv-Skalen Generalisiertes Vertrauen messen, ein erster empirischer Beweis zu ergeben. Mit anderen Worten: Sowohl bei der *Trust in People-Scale* als auch bei den Objektiv-Skalen scheint „Generalisiertes Vertrauen“ die Zieldimension zu sein. Diese Interpretation wird auch dadurch gestützt, dass sich bei Faktorenanalyse höherer Ordnung für die Objektiv-Skalen eine einzelne übergeordnete Dimension herauskristallisiert bzw. die Faktoren erster Ordnung stark miteinander korreliert sind.

Ausgehend von den beschriebenen Analyseergebnissen kann davon ausgegangen werden, dass es sich bei den Objektiv-Skalen um Messinstrumente handelt, die eine ausreichende Reliabilität besitzen. Mit anderen Worten: Die Messgenauigkeit der Objektiv-Skalen fällt mindestens „akzeptabel“ aus und erreicht das gleiche Niveau wie bei anderen Vertrauensskalen. Zudem scheint es sich bei beiden Objektiv-Skalen um verhältnismäßig homogene Tests zu handeln, die im Wesentlichen eine einzelne Dimension abbilden, die aufgrund theoretischer Vorüberlegungen als „Generalisiertes Vertrauen“ bezeichnet wurde. Die Ergebnisse der faktorenanalytischen Untersuchung konnten außerdem einen ersten empirischen Beleg dafür generieren, dass die Objektiv-Skalen die Zieldimension

„Generalisiertes Vertrauen“ tatsächlich abbilden. Dieser Eindruck konnte durch die Ergebnisse der Validitätsprüfung noch untermauert werden.

8.2 Validität der Objektiv-Skalen

Die Validität der Objektiv-Skalen wurde sowohl im Kontext der Pilotstudie als auch im Kontext der Hauptstudie geprüft. Allerdings ergaben sich hinsichtlich der angewandten Prüfverfahren Unterschiede zwischen den beiden Studien. So ging die Inhaltsvalidität in beiden Untersuchungen aus der theoriegeleiteten Operationalisierung der Zieldimension Generalisiertes Vertrauens bzw. der theoretischen Aufarbeitung dieses Konstrukts hervor. Ebenso wurde in beiden Studien eine Prüfung der Augenscheinvalidität und der Kriteriumsvalidität durchgeführt. Während allerdings das Prüfverfahren für die Augenscheinvalidität in Pilot- und Hauptstudie äquivalent war, wurden zur Bestimmung der Kriteriumsvalidität und im Speziellen der konkurrenten Validität bzw. Übereinstimmungsvalidität allerdings jeweils unterschiedliche externe Kriterien herangezogen. In der Pilotstudie waren dies subjektives Wohlbefinden, der sozioökonomische Status und die geographische Mobilität bzw. Stabilität des sozialen Umfeldes. In der Hauptstudie wurden das subjektive Wohlbefinden, das Vertrauen in Implementationsinstitutionen, die Eingebundenheit in informelle Netzwerke und die Mitgliedschaft in formellen Netzwerken verwendet. Die Konstruktvalidität wurde ausschließlich in der Hauptstudie überprüft. Aus forschungspraktischen und -ökonomischen Gründen wurde hierzu die *Trust in People-Scale* als Parallelmessung mit gleichem Gültigkeitsbereich verwendet. So sollten die Objektiv-Skalen ohnehin mit dem Standardinstrument zur Vertrauensmessung kontrastiert werden, entsprechend wurde die *Trust in People-Scale* auch zur Validierung genutzt.

Die erste Prüfung der Validität im Rahmen der Pilotstudie unterliegt zusätzlich gewissen Einschränkungen, die sich daraus ergeben, dass die Objektiv-Skalen dort erst entwickelt werden sollten. So liegt es in der Natur der Sache, dass die zur Bestimmung der Übereinstimmungsvalidität herangezogenen externen Kriterien in Abhängigkeit zur zu validierenden Zieldimension stehen sollten. Es ist in diesem Zusammenhang davon auszugehen, dass die Antworten der Probanden bezüglich dieser externen Kriterien unter dem Eindruck der 42 Items umfassenden Rohskala gegeben wurde. Die konkurrente Validität wird jedoch nur für die resultierende Objektiv-Skala bzw. Objektiv-Kurzskala bestimmt. Beide Skalen enthalten jedoch nur noch 14 bzw. 7 Items der ursprünglichen Rohskala. Es können daher bei der Bestimmung der Kriteriumsvalidität Verzerrungen nicht ausgeschlossen werden.

In der Hauptstudie werden den Probanden nur die 14 bzw. 7 Items umfassenden Objektiv-Skalen, die aus der Pilotstudie resultieren, zur Bearbeitung vorlegt. Entsprechend ist obengenannter Effekt hier eher nicht zu erwarten. Allerdings ist zu beachten, dass die Objektiv-Kurzskala in die Objektiv-Skala eingebettet ist, um zu vermeiden, dass einzelne Items doppelt beantwortet werden müssen. Strenggenommen könnte sich dadurch allerdings eine ähnlich gelagerte Einschränkung hinsichtlich der Resultate für die Kriteriumsvalidität der Objektiv-Kurzskala ergeben, wie sie weiter oben für die Pilotstudie beschrieben wurde.

Die Inhaltsvalidität der in den Objektiv-Skalen zusammengefassten Items bezüglich der Zieldimension „Generalisiertes Vertrauen“ ergibt sich aus deren Ableitung entlang theoretischer Vorüberlegungen. So stützt sich die hier vollzogene Operationalisierung Generalisierten Vertrauens auf die detaillierte theoretische Aufarbeitung des Zielkonstrukts und verwandter

8. Fazit

Aspekte bzw. Korrelate. Wie bereits beschrieben, war diese ausführliche und differenzierte Erörterung Generalisierten Vertrauens notwendig, um subtile Items formulieren zu können, die einen indirekten Zugang zur Zieldimension ermöglichen, ohne den Probanden einen unmittelbaren oder zu offensichtlichen Hinweis auf die Messintention der Objektiv-Skalen zu geben. Die Rohskala setzte sich dabei zu Beginn aus insgesamt 10 Faktoren (*Optimismus, Subjektive Schicksalskontrolle/Selbstkontrolle, Gemeinsames Schicksal/Gemeinsame Werte/Ähnlichkeit, Misanthropie, Vertrauensbereitschaft, Hilfsbereitschaft, Fairness Formelle/Informelle Netzwerke, Institutionen/Soziale Ordnung, Heterogenität/Diversität*) verteilt auf 42 Items zusammen, die entweder einzelne Facetten, Effekte, Konsequenzen, Voraussetzungen oder Begleiterscheinungen Generalisierten Vertrauens widerspiegeln. Die Elimination von Items im Zuge von Item- und Reliabilitätsanalyse reduzierte nicht nur die Faktorenzahl, sondern verschob auch die inhaltliche Schwerpunktsetzung der resultierenden Dimensionen der Objektiv-Skalen zu Gunsten einer Faktorstruktur, die der *Trust in People-Scale* ähnelte. Für die Objektiv-Skala ergaben sich in erster Instanz die Faktoren *Vertrauensbereitschaft, Misanthropie, Fairness* und *Hilfsbereitschaft*, während die Faktoren erster Ordnung bei der Objektiv-Kurzskala mit *Vertrauensbereitschaft* und *Fairness* benannt wurden. Alle diese Dimensionen waren bereits Bestandteil der Rohskala und ergeben sich somit aus der theoriegeleiteten Operationalisierung des Zielkonstrukts „Generalisiertes Vertrauen“. Allerdings entspricht die Itemzuordnung zu den resultierenden Faktoren der Objektiv-Skalen aufgrund der Itemreduktion nicht mehr der für die Faktorstruktur der Rohskala aufgelegten Zuweisung. Mit anderen Worten: In den Faktoren der Objektiv-Skalen sind teilweise andere Items enthalten als in den Faktoren der Rohskala mit gleichem Namen.

Es kann somit davon ausgegangen werden, dass die zu den Objektiv-Skalen zusammengefassten Items eine Abbildung des theoretischen Konstruktes „Generalisiertes Vertrauen“ ermöglichen bzw. eine präzise Messung der angestrebten Zieldimension zulassen. Mit anderen Worten: Für die Objektiv-Skalen ist Inhaltsvalidität gegeben.

Die Augenscheinvalidität der verwendeten Items bezüglich der Zieldimension „Generalisiertes Vertrauen“ spielte für die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen eine große Rolle. Während in der Regel bei der Erstellung von Skalen eine mit der Messintention übereinstimmende Augenscheinvalidität angestrebt wird, damit die Probanden ein unmittelbares Bewusstsein für den Zusammenhang von Testaufgaben und Zieldimension entwickeln, war es bei der Erstellung der Objektiv-Skalen das Ziel, Skalen zu erstellen, die eine geringe Augenscheinvalidität hinsichtlich der Zieldimension aufwiesen bzw. keine mit der Messintention übereinstimmende Augenscheinvalidität besaßen. Die erfolgreiche Verschleierung der Messintention stellte eine der Konstruktionsvorgaben für die Hybridklasse dar und sollte eine Reduktion reaktiver Effekte und im Speziellen sozialerwünschten Antwortverhaltens bewirken. Die hintergründige Idee war, dass die Probanden bei Unkenntnis der Messintention nicht zu zielgerichteten bzw. systematischen Antwortverfälschungen befähigt sind.

In der Pilotstudie gaben 20,5% der Probanden Vertrauen als mutmaßliche Zieldimension der Objektiv-Skalen an, während 79,5% ein von Vertrauen verschiedenes Konstrukt (z.B. Weltbild/Menschenbild, Normen und Werte, Gesellschaftliche Kohäsion, Sozialverhalten, Lebenszufriedenheit) als Zieldimension vermuteten. Aus der Berechnung der 99%-Konfidenzintervalle für diese Anteilswerte ging außerdem hervor, dass sich die Wertebereiche der Anteile nicht überschneiden bzw. sich der Anteil von Probanden, die die

8. Fazit

Messintention identifizieren konnten, stark vom Anteil der Probanden, die die Messintention nicht identifizieren konnten, unterscheidet.

Im Rahmen des Survey-Experimentes der Hauptstudie konnten voneinander unabhängige Prüfungen der Verschleierung der Messintention in den *Versuchsgruppen 1, 2 und 3* bzw. in allen Versuchsgruppen, die eine Abfrage Generalisierten Vertrauens per Objektiv-Skalen vornahmen, getätigt werden. Dabei konnten für die *Gruppen 1 und 3* das Ergebnis der Pilotstudie nahezu eins zu eins repliziert werden. In *Versuchsgruppe 1* nahmen 24% der Probanden Vertrauen als Zieldimension der Objektiv-Skalen an, während 76% der Probanden ein anderes Konstrukt als Zieldimension vermuteten. In *Versuchsgruppe 3* nahmen 22% der Probanden Vertrauen als Zieldimension an, während 78% der Probanden ein von Generalisiertem Vertrauen verschiedene Messintention bei den Objektiv-Skalen vermuteten. In beiden Fällen zeigen die 99%-Konfidenzintervalle der Anteilswerte an, dass sich der Anteil der Probanden, die die Messintention der Objektiv-Skalen erkannt haben, stark von dem Anteil der Probanden, die die Messintention der Objektiv-Skalen nicht erkannt haben, unterscheidet. In *Versuchsgruppe 2* liegen die Anteile von Identifikation und Nicht-Identifikation der Messintention näher beieinander. Hier gaben 48% der Probanden „Vertrauen“ als mutmaßliche Zieldimension der Objektiv-Skalen an, während 52% der Probanden ein anderes Konstrukt nannten. Aus der Berechnung der 99%-Konfidenzintervallen ging hervor, dass sich die Konfidenzintervalle der Anteilswerte überschneiden. Das bedeutet: Die Anteilswerte der Kategorien „Vertrauen“ und „anderes Konstrukt“ scheinen sich kaum bzw. nicht zu unterscheiden.

Verantwortlich für die unterschiedlichen Effekte in den verschiedenen Versuchsgruppen scheint die Präsentationsreihenfolge von Objektiv-Skalen und Standard-Skala im Zusammenspiel mit der Tatsache, dass die Messintention bei der Standard-Skala nicht verschleiert wird, zu sein. So werden in *Versuchsgruppe 1* die Objektiv-Skalen vor der Standard-Skala präsentiert und in *Versuchsgruppe 3* bearbeiten die Probanden ausschließlich die Objektiv-Skalen. In *Versuchsgruppe 2* wird dagegen die Standard-Skala vor den Objektiv-Skalen präsentiert. Entsprechend kann in *Versuchsgruppe 2* die Standard-Skala den Probanden einen Hinweis auf die Messintention der Objektiv-Skalen geben, während dies in den beiden übrigen Versuchsgruppen nicht möglich ist.

Im Schnitt konnten nur knapp ein Viertel der Probanden die Messintention der Objektiv-Skalen identifizieren, während knapp Dreiviertel der Probanden nicht dazu in der Lage waren, die Messintention der Objektiv-Skalen zu identifizieren. Die Verteilung aus *Versuchsgruppe 2* kann dagegen wohl als Ausreißer betrachtet werden, der auf die Bearbeitungsreihenfolge der Skalen zurückzuführen ist. Es ist vielmehr davon auszugehen, dass bei einer separaten Präsentation der Objektiv-Skalen, ohne dass umgebende Quellen Hinweise auf deren Zieldimension geben, in den meisten Fällen eine erfolgreiche Verschleierung der Messintention gewährleistet werden kann.

Insofern lässt sich für die Probanden per Augenschein eher kein unmittelbarer Zusammenhang zwischen den Items der Objektiv-Skalen und deren Zieldimension „Generalisiertes Vertrauen“ herstellen. Den Probanden ist es vielmehr in der Regel nicht ohne weiteres möglich, die Messintention der Objektiv-Items zu durchschauen. Das bedeutet: Die Objektiv-Items weisen hinsichtlich des theoretischen Konstruktes „Generalisiertes Vertrauen“ eine geringe Augenscheininvalidität auf bzw. es besteht tendenziell eher keine mit der Messintention der Objektiv-Skalen übereinstimmende *Face-Validity*.

Wie bereits beschrieben, wurde die Konstruktvalidität und im Speziellen die konvergente Validität der Objektiv-Skalen ausschließlich im Kontext der Hauptstudie geprüft. Als

vergleichbares Testinstrument mit gleichem oder ähnlichem Gültigkeitsbereich wie die Objektiv-Skalen wurde dabei die *Trust in People-Scale* verwendet, die hier als Kontrast zu den Objektiv-Skalen mit dem Begriff Standard-Skala bezeichnet wird und wurde. Die konvergente Validität der Objektiv-Skalen konnten im Rahmen des Survey-Experimentes nur in solchen Versuchsgruppen geprüft werden, in denen die Probanden sowohl die Objektiv-Skalen als auch die Standard-Skalen bearbeitet hatten. Entsprechend kamen nur die *Gruppen 1* und *2* infrage.

Die Korrelationen der Objektiv-Skalen mit der der Standard-Skala bewegen sich dabei in beiden Gruppen im Bereich von $.502 \leq r \leq .675$ und lagen damit oberhalb des Schwellenwertes für den Validitätskoeffizienten der konvergenten Validität von $r > .50$. Hieraus lässt sich zunächst ableiten, dass beide Objektiv-Skalen das gleiche theoretische Konstrukt erfassen wie die Standard-Skala. Zu bemerken ist allerdings, dass die Objektiv-Kurzskala jeweils stärker mit der Standard-Skala korreliert ist als die Objektiv-Skala. Dies legt die Vermutung nahe, dass die Objektiv-Kurzskala eine bessere Abbildung Generalisierten Vertrauens liefert als die Objektiv-Skala.

Das Radiusproblem der Vertrauensmessung geht im Wesentlichen davon aus, dass bei der Abbildung des Vertrauenslevels anhand der Standard-Skala Partikulares und Generalisiertes Vertrauen vermischt werden. Dieses Problem spielt auch für die Objektiv-Skalen eine Rolle, wie weiter unten berichtet wird. Strenggenommen bilden beide Skalentypen nur dann Generalisiertes Vertrauens ab, wenn die Probanden generalisierte Andere als Vertrauensreferenz heranziehen. Bezieht man das Radiusproblem in die Analyse der Konstruktvalidität ein, reduzieren sich die Korrelationen zwischen den Objektiv-Skalen und der Standard-Skala in *Gruppe 1* und *2*. Es ergeben nun Werte von $.422 \leq r \leq .590$, sodass die Werte der Validitätskoeffizienten teilweise unterhalb des Schwellenwertes liegen. Allerdings handelt es sich in beiden Versuchsgruppen um signifikante Zusammenhänge zwischen den Skalentypen. Außerdem ist zu beachten, dass sich aufgrund der Differenzierung entlang des Bezugsrahmens des Vertrauens, die in die Analyse eingehende Fallzahl im Vergleich zur vorangegangenen Analyse teilweise erheblich reduziert. So wird die Fallzahl in *Gruppe 1* mehr als halbiert und in *Gruppe 2* gehen 20 Fälle weniger als zuvor in die Analyse ein. Insbesondere in *Gruppe 2* sind die Skalentypen nun weniger stark korreliert, was vermuten lässt, dass auch die Präsentationsreihenfolge der Skalen einen Einfluss haben könnte. Möglicherweise gehen die Probanden in *Gruppe 2*, wo die Standard-Skala vor den Objektiv-Skalen bearbeitet wurde, davon aus, dass nicht zweimal das gleiche Konstrukt abgefragt wird. Entsprechend ergeben sich niedrigere Zusammenhangswerte zwischen den Skalentypen, da die Probanden bei den Objektiv-Skalen von einer im Vergleich zur Standard-Skala veränderten Zieldimension ausgehen. Diese Annahme ergibt sich daraus, dass die Messintention der Standard-Skala grundsätzlich transparent gestaltet war und sich daraus aufgrund der Präsentationsreihenfolge in *Versuchsgruppe 2* ein prospektiver Reihenfolgeeffekt ergeben könnte. Zu bemerken ist außerdem, dass bei Beachtung der Vertrauensreferenz die Schlussfolgerung, dass die Objektiv-Kurzskala eine im Vergleich zur Objektiv-Skala bessere Abbildung Generalisierten Vertrauens ermöglicht, weniger eindeutig ausfällt.

Insgesamt kann davon ausgegangen werden, dass die Items der Objektiv-Skalen ausreichend hoch mit den Items der *Trust in People-Scale* bzw. der Standard-Skala korreliert sind. Entsprechend kann für die Objektiv-Skalen Konstruktvalidität und im Besonderen konvergente Validität angenommen werden.

8. Fazit

Die Kriteriumsvalidität wurde in Pilot- und Hauptstudie anhand verschiedener externer Kriterien bestimmt. Die Auswahl der herangezogenen Kriterien war dabei entlang theoretischer Vorüberlegungen ausgerichtet, die sich aus der Aufarbeitung des empirischen und theoretischen Forschungsstandes bezüglich der Zieldimension „Generalisiertes Vertrauens“ ableiteten.

In der Pilotstudie waren dies subjektives Wohlbefinden, der sozioökonomische Status und die geographische Mobilität bzw. Stabilität des sozialen Umfeldes. Dabei ergaben sich zwischen den Messwerten der Objektiv-Skalen und den Messwerten der genannten Variablen Korrelationswerte im Bereich $-.018 \leq r \leq .134$. Damit liegen alle Validitätskoeffizienten unterhalb des Schwellenwertes für die Kriteriumsvalidität von $r > .20$. Den stärksten Effekt auf das Vertrauensniveau scheint dabei noch das subjektive Wohlbefinden („Zufriedenheit/Glück aktuell“ und „Lebenszufriedenheit allgemein“) der Probanden zu haben. Das durchschnittliche Monatseinkommen und die geographische Mobilität sowie die Stabilität des sozialen Umfeldes scheinen dagegen eher keinen Einfluss zu haben. Dies mag auf die Zusammensetzung der zugrundeliegenden Stichprobe zurückzuführen sein. So handelt es sich bei den Probanden ausnahmslos um Studierende. Diese soziale Gruppe verfügt in der Regel noch nicht über ein nennenswertes regelmäßiges Monatseinkommen aus eigener Arbeitsleistung. Vielmehr werden Studierende oft noch durch Leistungen der Eltern bzw. Verwandter oder entsprechender Studienförderung (z.B. BAföG) finanziert. Insofern liegt in vielen Fällen noch keine Eigenverantwortlichkeit für den Broterwerb vor bzw. es existiert ein soziales Fangnetz für die Studierenden. Es liegt daher die Vermutung nahe, dass das Monatseinkommen noch keinen besonderen Stellenwert für das Generalisierte Vertrauen von im Speziellen Studierenden besitzt, da ein entsprechender Ausfall von Dritten automatisch aufgefangen würde. Auch für den geringen Einfluss geographischer Mobilität und der Stabilität des sozialen Umfeldes könnten die Spezifika des Samples verantwortlich sein. Liegt der Studienort in (un-)mittelbarer Nähe des Heimatortes, so werden keine besonderen Anforderungen an die geographische Mobilität gestellt, da auch das gewohnte soziale Umfeld weitestgehend bestehen bleibt. Daneben zeichnen sich insbesondere Studierende durch eine selbstverständliche Bereitschaft zu geographischer Mobilität aus und auch die Bereitschaft, neue Kontakte einzugehen, ist in der Regel stark ausgeprägt. In beiden Fällen ist davon auszugehen, dass die genannten Variablen kaum eine Rolle für die Ausprägung des Generalisierten Vertrauens spielen. Vielmehr ist zu erwarten, dass geographische Mobilität und Stabilität des sozialen Umfeldes erst dann einen Effekt auf Generalisiertes Vertrauen haben, wenn es hierbei zu Brüchen kommt, auf die die Probanden nur beschränkten Einfluss haben. Dies scheint hier nicht gegeben, da in ersterem Fall keine Veränderung der Ausprägungen der Variablen stattfindet und in letzterem Fall eine Veränderung der Ausprägungen aktiv und willentlich herbeigeführt wird. Dabei spielen mit Sicherheit auch entsprechende Charakterdispositionen der Probanden eine Rolle.

Vor dem Hintergrund der zu Beginn dieses Abschnittes thematisierten Einschränkungen, der die Validitätsprüfungen in der Pilotstudie generell unterworfen sind und der Tatsache, dass die Validitätskoeffizienten durchweg unterhalb des Schwellenwertes rangieren, kann in der Pilotstudie keine Kriteriumsvalidität für die Objektiv-Skalen festgestellt werden.

In der Hauptstudie wurde ebenfalls das subjektive Wohlbefinden der Probanden als externes Kriterium herangezogen. Darüber hinaus wurden zusätzlich noch das Vertrauen in Implementationsinstitutionen, die Eingebundenheit in informelle Netzwerke und die Mitgliedschaft in formellen Netzwerken als Validierungskriterien für die Objektiv-Skalen

8. Fazit

verwendet. Die Prüfung der Kriteriumsvalidität der Objektiv-Skalen war dabei in den *Versuchsgruppen 1, 2 und 3* möglich, während in *Versuchsgruppe 4* nur die Standard-Skala präsentiert wurde.

Über die Gruppengrenzen hinweg hängen insbesondere das Vertrauen in Implementationsinstitutionen und die Eingebundenheit in informelle Netzwerke eng mit den Objektiv-Skalen zusammen. In *Gruppe 1* ergeben sich hier signifikante Korrelationen im Bereich $.295 \leq r \leq .409$. In *Gruppe 2* wird der Wertebereich $.155 \leq r \leq .278$ überspannt und in *Gruppe 3* ergeben sich zwischen den Objektiv-Skalen und den genannten externen Kriterien Korrelationswerte von $.175 \leq r \leq .226$. Die Validitätskoeffizienten nehmen dabei für beide Objektiv-Skalen mehrheitlich Werte oberhalb der Schwelle von $r > .20$ an. Es gilt somit: Je größer das Vertrauen in Implementationsinstitutionen (z.B. Judikative, Administration) und je stärker man in informelle Netzwerke (z.B. Familie, Freundschaften) eingebunden ist, desto höher fällt das Niveau Generalisierten Vertrauens aus und umgekehrt. Beide Beziehungen stehen im Einklang mit den zugehörigen theoretischen Vorüberlegungen. So wurde davon ausgegangen, dass politische Institutionen durch die Vorgabe und Gewährleistung einer gesellschaftsübergreifenden normativen Rahmung des erlaubten Handelns soziale Ordnung festigen und bestätigen. Politische Institutionen geben eine allgemeingültige Rechtsordnung vor und setzen diese durch. Auf diese Weise kann das zukünftige Verhalten der Mitbürger abgeschätzt werden und etwaiges Fehlverhalten würde sanktioniert werden. Die Folge ist ein wechselseitiges Vertrauen darauf, dass der spezifische, aber auch generalisierte Andere im Rahmen der institutionell vorgegebenen Regeln handeln wird und aus Interaktionen mit diesen Personen keine Nachteile für einen selbst entstehen.

Kontakt und Interaktion mit anderen Menschen, z.B. im Rahmen von Freundschaften, ermöglicht die Erfahrung wechselseitiger Vertrauenswürdigkeit und gegenseitigen Vertrauens. Dies reduziert die Bedrohlichkeit und erhöht die Attraktivität von Kooperationsbeziehungen. Kooperation wird vermehrt als lohnenswerte Aktivität angesehen, die Vorteile für eigene Vorgaben verspricht und erbringt. Positive Erfahrungen werden dann auf andere auch fremde Menschen übertragen, sodass sich mit der Zeit ein Generalisiertes Vertrauens entwickelt.

Der Zusammenhang zwischen subjektivem Wohlbefinden bzw. dessen Subkomponenten aktuelle Lebenszufriedenheit und allgemeine Lebenszufriedenheit und den Messwerten der Objektiv-Skalen fällt über die Gruppen betrachtet weniger eindeutig aus. Während in den *Gruppen 1 und 3* die Korrelationswerte entsprechend der Erwartungen durchweg ein positives Vorzeichen aufweisen, ergeben sich in *Gruppe 2* zwischen den Messwerten der Objektiv-Skalen und der aktuellen Lebenszufriedenheit negative Zusammenhänge. Das bedeutet: Entgegen der theoretischen Annahmen sinkt hier das Niveau Generalisierten Vertrauens mit dem Anstieg der aktuellen Lebenszufriedenheit und umgekehrt. Abhängig von der Richtung des Zusammenhangs nehmen die Korrelationen über alle Gruppen hinweg Werte von $.004 \leq r \leq .222$ bzw. $-.014 \leq r \leq -.161$ an, wobei die Validitätskoeffizienten mehrheitlich unterhalb des Schwellenwertes von $r > .20$ rangieren.

Lebenszufriedenheit macht Menschen attraktiver für andere, sodass sich für diese mehr Gelegenheit zu Kooperations- und Interaktionsbeziehungen ergeben. Dies hängt nicht zuletzt damit zusammen, dass zufriedene Menschen eher zu einer positiven Lebenseinstellung und einer optimistischen Weltwahrnehmung neigen als Menschen, die mit ihrem Leben unzufrieden sind. Positive Erfahrungen mit anderen Menschen bewirken dann, dass Menschen im Allgemeinen vertrauenswürdiger erscheinen und sich das Niveau Generalisierten Vertrauens steigert.

Die Variable „Mitgliedschaft in formellen Netzwerken“ ist eher schwach mit den Messwerten der Objektiv-Skalen korreliert. Zudem weist sie entgegen der theoretischen Annahmen in *Versuchsgruppe 2* einen negativen Zusammenhang mit den Objektiv-Skalen auf. Abhängig von der Richtung des Zusammenhangs schwanken die Werte der Korrelationen in den Gruppen zwischen $.000 \leq r \leq .126$ oder $-.062 \leq r \leq -.181$ und liegen damit durchweg unterhalb des Schwellenwertes für die Kriteriumsvalidität von $r > .20$.

Es wurde angenommen, dass innerhalb formeller sozialer Netzwerke, wie Parteien, Vereinen etc., sogenannte *civic virtues of trust* vermittelt werden, die Heuristiken zur Vergabe von Vertrauen entsprechen. Gleichzeitig lehrt das Engagement in solchen Vereinigungen eine Sensibilität und Aufmerksamkeit für Gemeinschaftsinteressen und Gemeinschaftsgüter. In der Folge wird beides auch auf den gesamtgesellschaftlichen Kontext übertragen, sodass nicht mehr nur Mitglieder des Netzwerkes, sondern auch Fremden vertraut wird.

In der Hauptstudie kann ausschließlich anhand der externen Kriterien „Vertrauen in Implementationsinstitutionen“ und „Eingebundenheit in informelle Netzwerke“ Kriteriumsvalidität für die Objektiv-Skalen abgeleitet werden. Die Validitätskoeffizienten liegen für die übrigen Kriterien „Subjektives Wohlbefinden“ und „Mitgliedschaft in formellen Netzwerken“ jeweils unterhalb des notwendigen Schwellenwertes oder weisen der vermuteten Zusammenhangsrichtung entgegengesetzte Vorzeichen auf. Insofern kann anhand dieser Variablen keine Kriteriumsvalidität für die Objektiv-Skalen abgeleitet werden.

Auch über Pilot- und Hauptstudie hinweg betrachtet lassen sich ausschließlich anhand des Vertrauens in Implementationsinstitutionen und der Eingebundenheit in informelle soziale Netzwerke Rückschlüsse auf das Niveau des Generalisierten Vertrauens ziehen. Auch wenn die übrigen externen Kriterien keinen nachhaltigen Beitrag zur Validitätsprüfung leisten, lässt sich insofern durchaus Kriteriumsvalidität für die Objektiv-Skalen feststellen.

Aus den vorangegangenen Analyseergebnissen lässt sich nun ableiten, dass die Objektiv-Skalen Inhaltsvalidität besitzen und Konstruktvalidität sowie Kriteriumsvalidität aufweisen. Ferner konnte eine geringe Augenscheinvalidität der Skalenitems hinsichtlich der Zieldimension realisiert werden, sodass an dieser Stelle die Designvorgaben für die Hybridklasse erfüllt werden.

8.3 Messwertunterschiede

Das Forschungsdesign des Survey-Experimentes, welches im Rahmen der Hauptstudie durchgeführt wurde, entsprach einer Kombination von *within-subject*- und *between-subject*-Design. Insgesamt wurden vier Versuchsgruppen realisiert, wobei die Probanden in den *Versuchsgruppen 1* und *2* jeweils beide Skalentypen (Objektiv-Skalen und Standard-Skala) bearbeiteten. Die Präsentationsreihenfolge der Skalen wurde dabei jedoch relativ zur jeweils anderen Gruppe invertiert. In den *Versuchsgruppen 3* und *4* bearbeiteten die Probanden jeweils nur einen Skalentyp. Diese experimentelle Anordnung wurde gewählt, um potentielle *Positions*- bzw. *Reihenfolge*-Effekte und *Carry-over*-Effekte kontrollieren bzw. identifizieren zu können.

Um auf Unterschiede im Antwortverhalten auf den verschiedenen Vertrauensskalen prüfen zu können, wurden zunächst die jeweiligen Skalensummenwerte berechnet. Deren Mittelwerte fanden dann Eingang in die Signifikanzprüfungen. Da mit jeweils 50 Fällen in den *Versuchsgruppen 1* und *2* und jeweils 52 Fällen in den *Versuchsgruppen 3* und *4* nur verhältnismäßig kleine Stichproben verwirklicht werden konnten, wurden zur wechselseitigen

Validierung der Ergebnisse bei der Prüfung auf überzufällige Differenzen zwischen den Antworten auf den verschiedenen Skalentypen sowohl parametrische als auch nonparametrische Verfahren eingesetzt.

Um zu prüfen, ob sich aus der experimentellen Situation und im Speziellen aus der Präsentationsreihenfolge der Skalen Unterschiede im Antwortverhalten ergeben, wurde zunächst geprüft, ob sich die Antworten auf der Standard-Skala über die Gruppengrenzen hinweg wechselseitig voneinander unterscheiden und ob sich die Antworten auf den Objektiv-Skalen über die Gruppengrenzen hinweg wechselseitig unterscheiden. Hierbei konnten keine überzufälligen Abweichungen im Antwortverhalten festgestellt werden.

Entsprechend wird angenommen, dass die Präsentationsreihenfolge der Skalen eher keinen Einfluss auf das Antwortverhalten der Probanden hatte, sondern etwaige Abweichungen vornehmlich auf die spezifischen Charakteristika der Skalentypen zurückzuführen sind.

Um zu prüfen, ob sich Unterschiede im Antwortverhalten auf den verschiedenen Skalentypen ergeben, wurden die Objektiv-Skalen mit der Standard-Skala verglichen. Entsprechend des *within-subject*-Designs wurden dabei die Skalentypen innerhalb der *Versuchsgruppen 1* und *2* kontrastiert. Dem *between-subject*-Design folgend wurden außerdem zusätzlich die Skalen aus den *Versuchsgruppen 3* und *4* gegenübergestellt. Zusätzlich konnten einige der Einzelmessungen auch als *One-Shot-Case-Study* bzw. *ex-post-facto-Design* interpretiert werden, sodass sich zwischen den Gruppen weitere Versuchsanordnungen im *between-subject*-Design ergaben.

Die Analyse erfolgte auf drei verschiedenen Differenzierungsstufen. Auf der ersten Stufe gingen alle gültigen Fälle in die Signifikanzprüfungen ein. Auf der zweiten Stufe wurde kontrolliert, ob die Identifikation der Messintention der Objektiv-Skalen als Moderator für potentielle Messwertunterschiede fungiert. Auf der dritten Stufe wurde zusätzlich zur Identifikation der Messintention auch die Vertrauensreferenz kontrolliert, um dem Radiusproblem Rechnung zu tragen.

Auf der ersten Differenzierungsstufe ergeben sich bei der Kontrastierung der Objektiv-Skalen und der Standard-Skala sowohl im *within-subject*- als auch im *between-subject*-Design vornehmlich signifikante Unterschiede. Dabei bestätigen sich die Resultate der parametrischen und nonparametrischen Signifikanzprüfungen wechselseitig weitestgehend. Unabhängig von der Signifikanz des Unterschiedes ergibt sich auf den Objektiv-Skalen relativ zur Standard-Skala grundsätzlich ein niedrigeres Vertrauenslevel. Die Probanden stellen sich auf der Standard-Skala somit signifikant vertrauensvoller dar als auf den Objektiv-Skalen. Dabei werden für den Kontrast von Objektiv- und Standard-Skala bei parametrischer Prüfung Teststärken von $.233 \leq 1 - \beta \leq .986$ und bei nonparametrischer Prüfung Teststärken von $.094 \leq 1 - \beta \leq .922$ erreicht. Für den Kontrast von Objektiv-Kurzskala und Standard-Skala werden entsprechend Teststärken von $.132 \leq 1 - \beta \leq .850$ und $.072 \leq 1 - \beta \leq .741$ erreicht. Hierbei ist zunächst zu beachten, dass die nicht-signifikanten Resultate grundsätzlich Teststärken unterhalb des notwendigen Schwellenwertes von $1 - \beta = .80$ und dementsprechend β -Fehlerwahrscheinlichkeiten oberhalb des Schwellenwertes von $\beta = .20$ aufwiesen. Es ist somit anzunehmen, dass es sich um falsch nicht-signifikante Ergebnisse handelt bzw. das Testergebnis nicht eindeutig interpretiert werden kann. Gleichzeitig erreichen aber auch die Teststärken einiger signifikanter Ergebnisse nicht das notwendige Niveau. Dies ist teilweise damit zu erklären, dass das 1%-Signifikanzniveau anstatt des 5%-Signifikanzniveaus angelegt werden konnte. In der Folge sank die Teststärke, lag aber bei $\alpha = .05$ oberhalb des Schwellenwertes.

8. Fazit

Insofern stellen solche Resultate keine falsch signifikanten Ergebnisse dar. Daneben existieren allerdings auch Fälle, bei denen die Teststärke auch bei $\alpha = .05$ unterhalb des Schwellenwertes lag. Hier ist die Gefahr eines α -Fehlers gegeben.

Die Effektstärken⁹² schwanken relativ stark in Abhängigkeit vom betrachteten Kontrast. Im Rahmen der parametrischen Analyse ergaben sich für die Gegenüberstellung von Objektiv- und Standard-Skala Werte von $.243 \leq g \leq .698$ und für den Unterschied zwischen Objektiv-Kurzskala und Standard-Skala ergaben sich Werte von $.164 \leq g \leq .489$. Bei der nonparametrischen Analyse konnten bei der Kontrastierung von Objektiv- und Standard-Skala Effektstärken von $.124 \leq \varphi \leq .555$ und bei der Kontrastierung von Objektiv-Kurzskala und Standard-Skala Effektstärken von $.089 \leq \varphi \leq .443$ beobachtet werden. Die Unterschiede im Antwortverhalten haben somit keine bis eine starke praktische Relevanz. Dabei werden jeweils die höchsten und niedrigsten Effekte in den *Versuchsgruppen 1* und *2* bzw. innerhalb des *within-subject*-Designs erreicht. Auffällig ist, dass die Effektstärken und damit die Unterschiede im Antwortverhalten insbesondere in *Versuchsgruppe 2* sehr niedrig ausfallen. Möglicherweise ergibt sich hier aus der Präsentationsreihenfolge der Vertrauensskalen ein Hinweis auf die Messintention der Objektiv-Skalen, sodass das Antwortverhalten konsistenter und damit der Unterschied geringer ausfällt. Außerdem zeigte sich, dass die Effektstärken für die Kontraste von Objektiv-Kurzskala und Standard-Skala im direkten Vergleich mit den Effektstärken für die Kontraste von Objektiv- und Standard-Skala jedes Mal niedriger ausfielen. Dies deutet daraufhin, dass das Antwortverhalten auf der Objektiv-Kurzskala eine größere Ähnlichkeit zum Antwortverhalten auf der Standard-Skala aufweist als das Antwortverhalten auf der Objektiv-Skala. Die Objektiv-Kurzskala scheint demnach der Standard-Skala ähnlicher zu sein, als die Objektiv-Skala, was schon auf Grundlage des Befundes zur Konstruktvalidität zu vermuten war.

Für die erste Differenzierungsstufe gilt somit, dass sich die gemittelten Skalensummenwerte der Objektiv-Skalen tendenziell signifikant von den gemittelten Skalensummenwerten der Standard-Skala unterscheiden. Auf den Objektiv-Skalen geben die Probanden im Durchschnitt ein anderes Vertrauensniveau an als auf der Standard-Skala. Dabei ist der Kontrast dergestalt, dass die Probanden auf den Objektiv-Skalen grundsätzlich ein niedrigeres Vertrauensniveau erreichen als auf der Standard-Skala. Aus den nicht-signifikanten Ergebnisse wird deutlich, dass die Objektiv-Skalen und die Standard-Skala in letzter Instanz mindestens äquivalente Messverfahren darstellen.

Auf der zweiten Differenzierungsstufe, d.h. bei Kontrolle der Identifikation der Messintention der Objektiv-Skalen, lässt sich zunächst beobachten, dass sich die Ergebnisse der parametrischen und der nonparametrischen Signifikanzprüfung für die Probandengruppe, die die Messintention der Objektiv-Skalen nicht identifizieren konnte (Gruppe B), wechselseitig bestätigen. Für die Probandengruppe, die die Messintention der Objektiv-Skalen identifizieren konnte, lassen sich dagegen die Ergebnisse von parametrischer und nonparametrischer Signifikanzprüfung nicht immer wechselseitig replizieren. Dies stellt einen Hinweis darauf dar,

⁹² Effektstärke bzw. Effektgröße: Hedges g bzw. Cohens d_s , mit g bzw. $d_s = \frac{|x_1 - x_2|}{\sigma_d} = \frac{|\bar{x}_d|}{\sigma_d}$, $d = .20 \rightarrow$ kleiner Effekt, $d = .50 \rightarrow$ mittlerer Effekt, $d = .80 \rightarrow$ großer Effekt (Eid/Gollwitzer/Schmitt (2013), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

Effektstärke bzw. Effektgröße: $\varphi = \sqrt{\frac{z^2}{N}} = \frac{z}{\sqrt{N}}$; mit $\omega = .10$ (kleiner Effekt), $\omega = .30$ (moderater Effekt), $\omega = .50$ (starker Effekt) (Bühner/Ziegler (2009), S. 269f; Cohen (1988), S. 223ff)

dass die Ergebnisse in Gruppe B trotz eventuell niedriger Teststärken valide sind, während den Ergebnissen in Gruppe A mit größerer Skepsis begegnet werden sollte.

Im Rahmen des *within-subject*- und des *between-subject*-Designs ergeben sich in Gruppe B sowohl bei der parametrischen als auch bei der nonparametrischen Signifikanzprüfung vornehmlich signifikante Unterschiede zwischen den Messwerten der Objektiv-Skalen und den Messwerten der Standard-Skala. Insgesamt stellen sich bei 8 von 16 Messwertpaarungen signifikante Unterschiede ein, während bei 4 von 16 Messwertpaarungen keine überzufällige Differenz auftritt. Dabei wird auf den Objektiv-Skalen immer ein niedrigeres Vertrauensniveau angegeben als auf der Standard-Skala. Das bedeutet: Die Objektiv-Skalen und die Standard-Skala stellen bei erfolgreicher Verschleierung der Messintention der Objektiv-Skalen mindestens äquivalente Messmethoden dar, in den meisten Fällen jedoch überschätzt die Standard-Skala das Vertrauensniveau relativ zu den Objektiv-Skalen bzw. die Objektiv-Skalen unterschätzen das Vertrauensniveau relativ zur Standard-Skala.

Die parametrischen Prüfungen erreichen dabei Teststärken von $.142 \leq 1 - \beta \leq .947$, während die verteilungsfreien Prüfungen Teststärken von $.073 \leq 1 - \beta \leq .817$ aufweisen. Alle nicht-signifikanten Ergebnisse weisen Teststärken unterhalb von $1 - \beta = .80$ und dementsprechend eine β -Fehlerwahrscheinlichkeiten von $\beta > .20$ auf. Es besteht somit die Gefahr, dass es sich hierbei um falsch nicht-signifikante Ergebnisse handeln könnte. Auch von den signifikanten Ergebnissen weist der Großteil Teststärken unterhalb des Schwellenwertes auf. Allerdings ergeben sich bei den parametrischen Signifikanzprüfungen Effektstärken von $.118 \leq g \leq .664$ und bei den nonparametrischen Signifikanzprüfungen Effektstärken von $.089 \leq \varphi \leq .558$. Alle signifikanten Messwertdifferenzen entsprachen dabei mindestens kleinen Effekten und können demnach als praktisch relevant eingestuft werden. Zudem waren die Ergebnisse von parametrischen und nonparametrischen Signifikanzprüfungen deckungsgleich.

In Gruppe A ergaben sich im Rahmen des *within-subject*- und *between-subject*-Design über die parametrischen und nonparametrischen Signifikanzprüfungen hinwegbetrachtet mehr nicht-signifikante als signifikante Unterschiede zwischen den Messwerten der Objektiv-Skalen und den Messwerten der Standard-Skala. Bei ersteren Prüfungen ergaben sich in 6 von 12 Messwertpaarungen signifikante Ergebnisse, während sich bei letzteren Prüfungen nur bei 2 von 12 Messwertpaarungen ein signifikantes Ergebnis einstellte. Die Resultate von parametrischer und nonparametrischer Signifikanzprüfung sind somit nicht für alle Messwertpaarungen deckungsgleich. Es gilt jedoch, dass bei Identifikation der Messintention der Objektiv-Skalen die Objektiv-Skalen und Standard-Skala somit vornehmlich gleichwertige Messverfahren darstellen.

Die parametrischen Prüfungen weisen Teststärken im Bereich von $.050 \leq 1 - \beta \leq .654$ auf und die nonparametrischen Prüfungen erreichen Teststärken von $.051 \leq 1 - \beta \leq .493$. Die Resultate können somit nicht statistisch abgesichert werden bzw. es besteht eine hohe β -Fehler-Wahrscheinlichkeit. Jedoch schwanken die Effektstärken der parametrischen Tests im Bereich $.014 \leq g \leq .664$ und die der nonparametrischen Tests im Bereich von $.089 \leq \varphi \leq .558$. Dabei ergaben sich insbesondere für die nicht-signifikanten Ergebnisse vornehmlich Effekte von $\omega < .10$. Das bedeutet: Die Differenz zwischen den Messwerten wird nicht als praktisch bedeutsam eingestuft. Insofern scheinen die nicht-signifikanten Resultate valide zu sein.

Aus der Analyse auf der zweiten Differenzierungsstufe geht hervor, dass sich die gemittelten Skalensummenwerte der Objektiv-Skalen insbesondere dann signifikant von den gemittelten Skalensummenwerten der Standard-Skala unterscheiden, wenn die Messintention der

Objektiv-Skalen erfolgreich verschleiert werden konnte. Dabei geben die Probanden auf den Objektiv-Skalen im Durchschnitt ein niedrigeres Vertrauensniveau an als auf der Standard-Skala. Wurde die Zieldimension der Objektiv-Skalen dagegen identifiziert, so unterscheiden sich die gemittelten Skalensummenwerte tendenziell nicht. Die Objektiv-Skalen und die Standard-Skala stellen dann äquivalente Messverfahren dar. Offensichtlich spielt die (Un-)Durchschaubarkeit der Messintention der Objektiv-Skalen bzw. die geringe Augenscheinvalidität der Items der Objektiv-Skalen bezüglich der Zieldimension eine Rolle für die Existenz von Messwertunterschieden zwischen den Objektiv-Skalen und der Standard-Skala.

Auf der dritten Differenzierungsebene wurde ebenfalls die Identifikation der Messintention kontrolliert, allerdings gingen nun nur noch Fälle in die Analyse ein, die generalisierte Andere als Vertrauensreferenz angeben hatten. Grund hierfür war das Radiusproblem der Vertrauensmessung, nach dem nur dann tatsächlich Generalisiertes Vertrauen erfasst wird, wenn die Probanden als Bezugsrahmen ihres Vertrauens die *outgroup* nennen. Dies hatte jedoch zur Folge, dass die Zahl der zu berücksichtigenden Fälle teilweise sehr stark absank. Die beobachteten Effekte können in einem solchen Fall auch auf das dann gesteigerte Gewicht von extremen Antworten zurückgehen.

Im Rahmen des *within-subject*-Design wurden ausschließlich nonparametrische Signifikanzprüfungen durchgeführt, da die Fallzahl für parametrische Verfahren nicht ausreichend war. Hierbei ergaben sich sowohl unter der Bedingung, dass die Probanden die Messintention der Objektiv-Skalen identifizieren konnten als auch unter der Bedingung, dass die Identifikation der Messintention der Objektiv-Skalen nicht gelang, signifikante und nicht-signifikante Ergebnisse. Insgesamt überwiegen jedoch Resultate, nach denen kein signifikanter Unterschied zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala besteht. Weder die signifikanten noch die nicht-signifikanten Ergebnisse können jedoch statistisch abgesichert werden. Es ergaben sich ausschließlich Teststärken von $.064 \leq 1 - \beta \leq .697$, sodass für die nicht-signifikanten Ergebnisse ein erhöhtes β -Fehler-Risiko besteht. Die Effektstärken der Unterschiede zwischen den Objektiv-Skalen und der Standard-Skala fielen in der Probandengruppe, die die Messintention der Objektiv-Skalen identifizieren konnte (Gruppe A: $.372 \leq \varphi \leq .866$), größer aus als in der Probandengruppe, die diese Messintention nicht identifizieren konnte (Gruppe B: $.092 \leq \varphi \leq .602$). Augenscheinlich fällt die Differenz zwischen den Messwerten der verschiedenen Vertrauensskalen somit in Gruppe A tendenziell größer aus als in Gruppe B. Hierbei ist allerdings zu beachten, dass in Gruppe A teilweise nur halb so viele Fälle Eingang fanden wie in Gruppe B. Es ist daher anzunehmen, dass extremes Antwortverhalten in Gruppe A voraussichtlich eher Messwertunterschiede bewirkt als in Gruppe B.

Im Rahmen des *between-subject*-Design wurden parametrische und nonparametrische Signifikanzprüfungen durchgeführt, deren Ergebnisse sich wechselseitig bestätigten. Allerdings gingen auch hier insbesondere in die Analysen von Gruppe A teilweise nur wenige Fälle ein. Extreme Antworten könnten in dieser Gruppe somit einen besonders großen Effekt haben.

Über alle Skalen hinweg und unabhängig von der Identifikation der Messintention der Objektiv-Skalen ergaben sich bei der Unterschiedsprüfung vornehmlich nicht-signifikante Ergebnisse, wobei in der Mehrzahl der Fälle auf den Objektiv-Skalen ein niedrigeres Vertrauensniveau angegeben wurde als auf der Standard-Skala. In den verbleibenden Fällen fielen die durchschnittlichen Skalensummenwerte exakt gleich groß aus. Signifikante

Unterschiede ergaben sich sowohl in Gruppe A als auch in Gruppe B nur zwischen den Messwerten von Objektiv- und Standard-Skala. Zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala ergaben sich dagegen sowohl in Gruppe A als auch in Gruppe B keine überzufälligen Abweichungen. Keines der Ergebnisse der Signifikanzprüfungen konnte statistisch abgesichert werden, da die Teststärken bei den parametrischen Verfahren mit $.050 \leq 1 - \beta \leq .596$ und bei den nonparametrischen Verfahren mit $.050 \leq 1 - \beta \leq .182$ weit unterhalb des Schwellenwertes von $1 - \beta = .80$ lagen. Für die nicht-signifikanten Ergebnisse geht hiermit eine hohe β -Fehler-Wahrscheinlichkeit bzw. die Gefahr von falsch nicht-signifikanten Ergebnissen einher. Die Effektstärken fallen, wie zu erwarten, bei den signifikanten Unterschieden am höchsten aus. Je nach angewandter Signifikanzprüfung schwanken sie jedoch verhältnismäßig extrem im Bereich von $0 \leq g \leq .711$ bzw. $.009 \leq \varphi \leq .287$. Dabei ergaben sich im Rahmen der parametrischen und nonparametrischen Prüfungen in Gruppe A überwiegend keine praktisch relevanten Differenzen zwischen den Messwerten der Vertrauensskalen, während die Effektstärken in Gruppe B durchgängig im Bereich für niedrige bis mittlere Effekte verortet waren. Trotz der vornehmlich nicht-signifikanten Messwertunterschiede zwischen den Objektiv-Skalen und der Standard-Skala können somit bei erfolgreicher Verschleierung der Messintention und generalisierten Anderen als Bezugsrahmen des Vertrauens die Differenzen zwischen den gemittelten Skalensummenwerten als praktisch bedeutsam eingestuft werden. Dagegen sind die Differenzen bei Identifikation der Messintention der Objektiv-Skalen größtenteils nicht praktisch bedeutsam.

Auf der dritten Differenzierungsstufe, d.h. bei Kontrolle der Identifikation der Messintention der Objektiv-Skalen und unter Berücksichtigung der Vertrauensreferenz „generalisierte Andere“, ergeben sich zwischen den gemittelten Skalensummenwerten der Objektiv-Skalen und der Standard-Skala mehrheitlich keine signifikanten Unterschiede. Die gefundenen überzufälligen Messwertunterschiede zwischen Objektiv- und Standard-Skala treten unabhängig von der Identifikation der Messintention der Objektiv-Skala auf und sind deshalb vermutlich eher auf die Unterschiede in der inhaltlichen Zusammensetzung der Skalen sowie der unterschiedlichen Skalenlängen und den daraus resultierenden verschiedenen Differenzierungsniveaus der Skalen bezüglich der Zieldimension „Generalisiertes Vertrauen“ zurückzuführen. Bei erfolgreicher Verschleierung der Messintention fallen jedoch die Differenzen zwischen den gemittelten Skalensummenwerten der Objektiv-Skalen und der Standard-Skala größer aus, als wenn die Messintention identifiziert werden konnte, und auf den Objektiv-Skalen wird ein niedrigeres Vertrauensniveau angegeben als auf der Standard-Skala. Dagegen ergeben sich bei aufgedeckter Zieldimension der Objektiv-Skalen verschiedene Relationen der gemittelten Skalensummenwerte. Bei Beachtung der Vertrauensreferenz stellen Objektiv-Skalen und Standard-Skala unabhängig von der Identifikation der Zieldimension mindestens äquivalente Messverfahren dar. Es ergeben sich allerdings auch hier Hinweise darauf, dass die (Un-)Durchschaubarkeit der Messintention der Objektiv-Skalen einen Beitrag zu den Messwertunterschieden zwischen Objektiv-Skalen und Standard-Skala leistet.

Insgesamt ist anzunehmen, dass sich die gemittelten Skalensummenwerte der Objektiv-Skalen tendenziell signifikant von den gemittelten Skalensummenwerten der Standard-Skala unterscheiden. Insbesondere die (Un-)Durchschaubarkeit der Messintention der Objektiv-Skalen scheint hierbei eine große Rolle zu spielen. Dies lässt sich daraus ableiten, dass bei erfolgreicher Verschleierung der Zieldimension „Generalisiertes Vertrauen“ auf den Objektiv-Skalen immer ein niedrigeres Vertrauensniveau erreicht wird als auf der Standard-Skala.

Konnte die Messintention dagegen identifiziert werden, ergeben sich auch andere Verhältnisse für die gemittelten Skalensummenwerte der verschiedenen Vertrauensskalen. Das heißt: Sie fallen exakt gleich aus oder auf der Standard-Skala wird ein niedrigeres Vertrauensniveau erreicht als auf den Objektiv-Skalen.

Daneben kristallisieren sich allerdings auch Hinweise darauf heraus, dass das Radiusproblem bzw. die Diffusion der Zieldimensionen „Partikulares und Generalisiertes Vertrauen“ einen Einfluss auf die Mittelwertunterschiede haben könnten. So ergeben sich bei Kontrolle der Vertrauensreferenz kaum signifikante Messwertunterschiede zwischen den verschiedenen Vertrauensskalen. Dies legt die Vermutung nahe, dass die zuvor gefundenen überzufälligen Differenzen zwischen den Messwerten der Vertrauensskalen auf die Vermischung zweier verschiedener Zieldimensionen zurückgehen. Dagegen lässt sich argumentieren, dass die Relation der gemittelten Skalensummenwerte von Objektiv-Skalen und Standard-Skala auf allen Differenzierungsstufen bei erfolgreicher Maskierung der Zieldimension der Objektiv-Skalen stabil bleibt und die Signifikanzprüfungen insbesondere bei Berücksichtigung der Vertrauensreferenz teilweise anhand vergleichsweise geringen Fallzahlen durchgeführt wurden. Außerdem zeigte ein Vergleich der Vertrauensniveaus von Probanden, die generalisierte Andere als Bezugsrahmen ihres Vertrauens angegeben hatten, mit dem Vertrauensniveau von Probanden, die spezifische Andere als Bezugsrahmen ihres Vertrauens angegeben hatten, keine überzufälligen Unterschiede. Das Niveau Partikularen und Generalisierten Vertrauens fällt somit gleich groß aus, sodass keine Verzerrungen der gemittelten Skalensummenwerte aufgrund der Durchmischung der beiden Zieldimensionen zu erwarten sind.

8.4 Einfluss reaktiver Effekte

Reaktive Effekte können verschiedene Formen annehmen. Im Rahmen dieser empirischen Untersuchung standen Verzerrungen des Antwortverhaltens infolge von sozialer Erwünschtheit im Fokus des Interesses. Dabei wurde sozialerwünschtes Antwortverhalten bezüglich der Zieldimension „Generalisiertes Vertrauen“ als *fake good* bzw. als Neigung zur Überschätzung des eigenen Vertrauensniveaus identifiziert. Aus den Ergebnissen der Mittelwertvergleiche von Objektiv-Skalen und Standard-Skala lässt sich ableiten, dass die Probanden ihr Vertrauensniveau auf der Standard-Skala in Relation zu ihren Angaben auf den Objektiv-Skalen tendenziell überschätzten. Dieses Ergebnis lässt sich nun dahingehend interpretieren, dass das per Standard-Skala ermittelte Vertrauensniveau aufgrund sozialerwünschten Antwortverhaltens zustande gekommen ist und das per Objektiv-Skalen ermittelte Vertrauensniveau eher der Wirklichkeit entspricht.

Der Einfluss sozialerwünschten Antwortverhaltens auf die Objektiv-Skalen wurde sowohl in der Pilotstudie als auch in der Hauptstudie anhand des Zusammenhangs der Vertrauensskalen mit sogenannten SD-Skalen oder „Lügenskalen“ untersucht. In der Pilotstudie wurde hierzu die *SDS-CM (Kurzsкала)*, eine aus der deutschen Übersetzung der *Marlowe-Crowne-Scale* hervorgehende Kurzsкала, verwendet, während in der Hauptstudie die *BIDR-Kurzsкала (Balanced Inventory of Desirable Responding)* zum Einsatz kam. In der Hauptstudie wurde außerdem noch ein Vergleich der Anfälligkeit von Objektiv-Skalen und Standard-Skala für sozialerwünschte Antworten realisiert und der Effekt sozialer Erwünschtheit auf die Differenz zwischen den gemittelten Skalensummenwerten analysiert.

8. Fazit

Die Analyse des Einflusses sozialer Erwünschtheit auf die Objektiv-Skalen im Rahmen der Pilotstudie ergab einen schwachen Zusammenhang zwischen SD-Skala und den Objektiv-Skalen. Es ist daher anzunehmen, dass die Messwerte der Objektiv-Skalen verhältnismäßig gering durch sozialerwünschtes Antwortverhalten verfälscht sind.

Die Korrelationen der Objektiv-Skalen und der Standard-Skala mit der SD-Skala wurden in der Hauptstudie, wie bereits die Mittelwertvergleiche, auf drei verschiedenen Differenzierungsstufen betrachtet. Auf der ersten Stufe gingen alle gültigen Fälle in die Korrelationsanalyse ein. Auf der zweiten Stufe wurde differenziert, ob die Probanden die Messintention der Objektiv-Skalen identifizieren konnten. Auf der dritten Stufe wurde zusätzlich zur Identifikation der Zieldimension auch noch die angegebene Vertrauensreferenz berücksichtigt.

Unter Einbezug aller gültigen Fälle in die Korrelationsanalyse ergab sich durch den Einfluss sozialerwünschten Antwortverhaltens kaum ein Unterschied in der Verzerrung der Messwerte zwischen den verschiedenen Skalentypen. In *Versuchsgruppe 1* korrelierte nur die Objektiv-Kurzskala weniger stark mit der SD-Skala als die Standard-Skala. Die Objektiv-Skala wies dagegen einen größeren Zusammenhang mit der SD-Skala auf als die Standard-Skala. Dabei ergaben sich über alle Skalentypen hinweg geringe bis mittlere Zusammenhänge mit der SD-Skala ($.10 \leq r \leq .50$). (vgl. Tab.9.3) In *Versuchsgruppe 2* korrelierten die Objektiv-Skalen und die Standard-Skala annähernd gleich stark mit der SD-Skala. Es ergaben sich maximal schwache Zusammenhänge ($.00 \leq r \leq .10$). (vgl. Tab. 9.3) In *Versuchsgruppe 3* und *4* korrelierten die Objektiv-Skalen und die Standard-Skala ebenfalls annähernd gleich stark mit der SD-Skala, wobei sich jeweils schwache Zusammenhänge ergaben ($.00 \leq r \leq .10$). (vgl. Tab. 9.1 bzw. Tab. 9.2)

Hierbei ist allerdings zu berücksichtigen, dass die geringe Verfälschung der Messwerte teilweise auf die wechselseitige Kompensation von Selbst- und Fremdtäuschungstendenz zurückzuführen ist. In *Versuchsgruppe 1* konnte kein Kompensationseffekt festgestellt werden. Die Korrelationen mit der SDE- und der IM-Komponente entsprechen weitestgehend den Korrelationen mit der gesamten BIDR-Kurzskala ($.00 \leq r \leq .30$). (vgl. Tab. 9.1) In *Versuchsgruppe 2* ist die Objektiv-Kurzskala am stärksten und die Objektiv-Skala am wenigsten von einer wechselseitigen Kompensation der Subkomponenten sozialer Erwünschtheit betroffen. Die Standard-Skala ist in der Mitte anzusiedeln. Die Neigung zur Selbsttäuschung bewirkt hier tendenziell ein Absinken der Messwerte, während die Neigung zur Fremdtäuschung einen Anstieg bewirkt. Dabei fällt der Einfluss der Fremdtäuschungstendenz ($.00 \leq r \leq .10$) jedoch schwächer als der der Selbsttäuschungstendenz ($.10 \leq r \leq .30$) aus. (vgl. Tab. 9.1) In *Versuchsgruppe 3* und *4* ergeben sich ebenfalls Kompensationseffekte von SDE- und IM-Komponente. Diese fallen insbesondere in *Versuchsgruppe 3*, d.h. bezüglich der Objektiv-Skalen, vergleichsweise stark aus. Die SDE-Tendenz bewirkt in beiden Versuchsgruppen ein Absinken des Vertrauensniveaus, während die IM-Tendenz ein Ansteigen des Vertrauensniveaus verursacht. Dies entspricht den Ergebnissen aus *Versuchsgruppe 2*. In *Versuchsgruppe 3* sind die Korrelationen der Objektiv-Skalen mit den Subkomponenten jeweils ähnlich stark ausgeprägt und entsprechen einem mittleren Zusammenhang (vgl. Tab. 9.1). In *Versuchsgruppe 4* fällt die Korrelation der Standard-Skala mit der SDE-Komponente etwas stärker aus als die Korrelation mit der IM-Komponente. Sie entsprechen dabei keinem bis zu einem schwachen Zusammenhang ($.00 \leq r \leq .30$) (vgl. Tab. 9.2).

Es ist nun anzunehmen, dass die Objektiv-Messungen ähnlich stark mit der SD-Skala korreliert sind wie die Standard-Messung. Für beide Skalentypen gilt, dass ihre Messwerte tendenziell wenig durch sozialerwünschtes Antwortverhalten verfälscht werden. Diese

geringe Verzerrung der Messwerte ist jedoch in beiden Fällen Resultat eines Kompensationseffektes der Subkomponenten (Selbst- vs. Fremdtäuschung) sozialer Erwünschtheit. Dieser Effekt scheint in der Tendenz für die Objektiv-Skalen etwas stärker ausgeprägt zu sein, was zur Folge hat, dass deren Messwerte insgesamt etwas weniger stark verzerrt sind als die Messwerte der Standard-Skala. Es bedeutet jedoch gleichzeitig auch, dass auf den Objektiv-Skalen eigentlich weniger aufrichtig geantwortet wird als auf der Standard-Skala.

Bei Berücksichtigung aller Fälle hat sozialerwünschtes Antwortverhalten keinen nachhaltigen Einfluss auf die Differenz zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala. In *Versuchsgruppe 1* ergeben sich kein bis ein schwacher Zusammenhang zwischen der Messwertdifferenz und der SD-Skala ($.00 \leq r \leq .30$) (vgl. Tab. 9.3). In *Versuchsgruppe 2* ergibt sich kein nennenswerter Zusammenhang von Messwertdifferenz und SD-Skala ($.00 \leq r \leq .10$) (vgl. Tab. 9.3).

Es ist somit zunächst davon auszugehen, dass der Unterschied zwischen den gemittelten Skalensummenwerten der Objektiv-Skalen und der Standard-Skala nicht auf Verfälschungen infolge sozialerwünschten Antwortverhaltens zurückzuführen ist.

8.4.1 Kontrolle von Augenscheinvalidität und Vertrauensreferenz

Die geringe Augenscheinvalidität der Items bezüglich der jeweiligen Zieldimension stellt eine der Schlüsseleigenschaften des Designs der Objektiv-Skalen dar und beschreibt die zentrale Maßnahme, um eine Reduktion von Messwertverzerrungen infolge reaktiver Effekte im Vergleich zu traditionellen Befragungstechniken herbeizuführen. So soll die erfolgreiche Verschleierung der Messintention zielgerichtete Verfälschungen des Antwortverhaltens wirksam verhindern oder zumindest eindämmen, indem mindestens eine Ambivalenz des Anforderungscharakters der Messung hervorgerufen wird. Durch die Uneindeutigkeit des Messziels wird es für die Probanden schwierig bis unmöglich zu bestimmen, welches Antwortverhalten einem besonders positiven bzw. besonders negativen Eindruck entspricht. Mit anderen Worten: Es lässt sich nicht eindeutig ermitteln, was einem sozialerwünschten Antwortverhalten entspricht.

In diesem Zusammenhang konnte zunächst gezeigt werden, dass die Messwerte der Objektiv-Skala bei erfolgreicher Verschleierung der Messintention in der Tendenz vergleichsweise weniger anfällig für sozialerwünschtes Antwortverhalten sind bzw. bei nachweislicher Intransparenz der Zieldimension weniger stark durch sozialerwünschtes Antwortverhalten verzerrt werden. Bezüglich der Objektiv-Kurzskala ließ sich zumindest eine Tendenz in die Richtung einer solchen Wirkung beobachten. Allerdings handelt es sich in beiden Fällen nicht um einen Effekt, der in allen Versuchsgruppen stabil ist bzw. in gleicher Art und Weise sichtbar wird.

8. Fazit

Tab. 8.1: Verfälschung der Messwerte der Objektiv-Skalen durch sozialerwünschtes Antwortverhaltens sowie Kompensation von SDE- und IM-Komponente

VG	Diff.- stufe	Relation von $ r_{OK;BIDR} $ und $ r_{OKS;BIDR} $	Wertebereich von $ r_{OK;BIDR} $ und $ r_{OKS;BIDR} $	Relation von $ r_{OK;SDE} $ und $ r_{OK;IM} $	Wertebereich von $ r_{OK;SDE} $ und $ r_{OK;IM} $	Relation von $ r_{OKS;SDE} $ und $ r_{OKS;IM} $	Wertebereich von $ r_{OKS;SDE} $ und $ r_{OKS;IM} $
1	1	$OK(+) > OKS(+)$	$.10 \leq r \leq .50$	$SDE(+) < IM(+)$	$.10 \leq r \leq .30$	$SDE(+) \approx IM(+)$	$.00 \leq r \leq .10$
	2a	$OK(+) > OKS(+)$	$.00 \leq r \leq .50$	$SDE(-) < IM(+)$	$.00 \leq r \leq .50$	$SDE(-) > IM(+)$	$.10 \leq r \leq .50$
	2b	$OK(+) \approx OKS(+)$	$.10 \leq r \leq .30$	$SDE(+) > IM(+)$	$.10 \leq r \leq .30$	$SDE(+) > IM(0)$	$.00 \leq r \leq .30$
	3a	$OK(+) > OKS(+)$	$.00 \leq r \leq .30$	$SDE(-) < IM(+)$	$.10 \leq r \leq .50$	$SDE(-) > IM(+)$	$.30 \leq r \leq .70$
	3b	$OK(+) > OKS(+)$	$.10 \leq r \leq .70$	$SDE(+) > IM(+)$	$.10 \leq r \leq .70$	$SDE(+) > IM(-)$	$.10 \leq r \leq .70$
2	1	$OK(-) \approx OKS(+)$	$.00 \leq r \leq .10$	$SDE(-) > IM(+)$	$.00 \leq r \leq .30$	$SDE(-) > IM(+)$	$.10 \leq r \leq .30$
	2a	$OK(-) < OKS(-)$	$.10 \leq r \leq .30$	$SDE(-) < IM(-)$	$.00 \leq r \leq .50$	$SDE(-) > IM(+)$	$.00 \leq r \leq .50$
	2b	$OK(+) < OKS(+)$	$.00 \leq r \leq .30$	$SDE(-) < IM(+)$	$.10 \leq r \leq .50$	$SDE(-) < IM(+)$	$.00 \leq r \leq .50$
	3a	$OK(-) \approx OKS(+)$	$.10 \leq r \leq .30$	$SDE(+) < IM(-)$	$.00 \leq r \leq .50$	$SDE(-) > IM(-)$	$.30 \leq r \leq .70$
	3b	$OK(-) < OKS(-)$	$.10 \leq r \leq .70$	$SDE(-) > IM(+)$	$.10 \leq r \leq .50$	$SDE(-) \approx IM(+)$	$.00 \leq r \leq .10$
3	1	$OK(+) \approx OKS(-)$	$.00 \leq r \leq .10$	$SDE(-) < IM(+)$	$.30 \leq r \leq .50$	$SDE(-) > IM(+)$	$.10 \leq r \leq .50$
	2a	$OK(+) > OKS(+)$	$.10 \leq r \leq .30$	$SDE(-) < IM(+)$	$.10 \leq r \leq .70$	$SDE(+) < IM(+)$	$.00 \leq r \leq .30$
	2b	$OK(+) \approx OKS(-)$	$.00 \leq r \leq .10$	$SDE(-) \approx IM(+)$	$.30 \leq r \leq .50$	$SDE(-) > IM(+)$	$.30 \leq r \leq .50$
	3a	$OK(+) > OKS(+)$	$.50 \leq r \leq .70$	$SDE(-) < IM(+)$	$.30 \leq r \leq 1.00$	$SDE(+) > IM(+)$	$.10 \leq r \leq .70$
	3b	$OK(+) > OKS(+)$	$.00 \leq r \leq .30$	$SDE(-) < IM(+)$	$.30 \leq r \leq .50$	$SDE(-) > IM(+)$	$.30 \leq r \leq .70$

VG: Versuchsgruppe

 $|r_{OK;BIDR}|$ bzw. (OK): Betrag der Korrelation der Objektiv-Skala mit der SD-Skala (hier: BIDR-Kurzskala) $|r_{OKS;BIDR}|$ bzw. (OKS): Betrag der Korrelation der Objektiv-Kurzskala mit der SD-Skala (hier: BIDR-Kurzskala) $|r_{OK;SDE}|$ bzw. SDE: Betrag der Korrelation der Objektiv-Skala mit der Selbsttäuschungskomponente (SDE) der BIDR-Kurzskala $|r_{OK;IM}|$ bzw. IM: Betrag der Korrelation der Objektiv-Skala mit der Fremdtäuschungskomponente (IM) der BIDR-Kurzskala $|r_{OKS;SDE}|$ bzw. SDE: Betrag der Korrelation der Objektiv-Kurzskala mit der Selbsttäuschungskomponente (SDE) der BIDR-Kurzskala $|r_{OKS;IM}|$ bzw. IM: Betrag der Korrelation der Objektiv-Kurzskala mit der Fremdtäuschungskomponente (IM) der BIDR-Kurzskala

(+: positiver Zusammenhang)

(-): negativer Zusammenhang

1. Differenzierungsstufe: Alle gültigen Fälle

2. Differenzierungsstufe: Unterscheidung nach Identifikation der Messintention der Objektiv-Skalen (a: Messintention identifiziert; b: Messintention verschleiert)

3. Differenzierungsstufe: Unterscheidung nach Identifikation der Messintention der Objektiv-Skalen und entlang der Vertrauensreferenz (a: Messintention identifiziert; b: Messintention verschleiert)

8. Fazit

Tab. 8.2: Verfälschung der Messwerte der Standard-Skala durch sozialerwünschtes Antwortverhaltens sowie Kompensation von SDE- und IM-Komponente

VG	Diff.-stufe	Vorzeichen von $r_{S,BIDR}$	Wertebereich von $ r_{S,BIDR} $	Relation von $ r_{S,SDE} $ und $ r_{S,IM} $	Wertebereich von $ r_{S,SDE} $ und $ r_{S,IM} $
1	1	$S(+)$	$.10 \leq r \leq .30$	$SDE(+) = IM(+)$	$.10 \leq r \leq .30$
	2a	$S(+)$	$.10 \leq r \leq .30$	$SDE(-) \approx IM(+)$	$.10 \leq r \leq .30$
	2b	$S(+)$	$.10 \leq r \leq .30$	$SDE(+) > IM(+)$	$.00 \leq r \leq .30$
	3a	$S(+)$	$.50 \leq r \leq .70$	$SDE(-) < IM(+)$	$.00 \leq r \leq .70$
	3b	$S(+)$	$.30 \leq r \leq .50$	$SDE(+) > IM(+)$	$.00 \leq r \leq .70$
2	1	$S(-)$	$.00 \leq r \leq .10$	$SDE(-) > IM(+)$	$.00 \leq r \leq .30$
	2a	$S(0)$	$.00 \leq r \leq .10$	$SDE(-) > IM(+)$	$.10 \leq r \leq .30$
	2b	$S(-)$	$.00 \leq r \leq .10$	$SDE(-) \approx IM(+)$	$.00 \leq r \leq .10$
	3a	$S(-)$	$.10 \leq r \leq .30$	$SDE(-) < IM(-)$	$.10 \leq r \leq .30$
	3b	$S(-)$	$.00 \leq r \leq .10$	$SDE(-) \approx IM(-)$	$.00 \leq r \leq .10$
4	1	$S(-)$	$.00 \leq r \leq .10$	$SDE(-) \approx IM(+)$	$.00 \leq r \leq .30$
	3c	$S(+)$	$.00 \leq r \leq .10$	$SDE(-) < IM(+)$	$.10 \leq r \leq .30$

VG: Versuchsgruppe

 $|r_{S,BIDR}|$ bzw. (S): Betrag der Korrelation der Standard-Skala mit der SD-Skala (hier: BIDR-Kurzskala) $|r_{S,SDE}|$ bzw. SDE: Betrag der Korrelation der Standard-Skala mit der Selbsttäuschungskomponente (SDE) der BIDR-Kurzskala $|r_{S,IM}|$ bzw. IM: Betrag der Korrelation der Standard-Skala mit der Fremdtäuschungskomponente (IM) der BIDR-Kurzskala

(+: positiver Zusammenhang)

(-): negativer Zusammenhang

1. Differenzierungsstufe: Alle gültigen Fälle

2. Differenzierungsstufe: Unterscheidung nach Identifikation der Messintention der Objektiv-Skalen (a: Messintention identifiziert; b: Messintention verschleiert)

3. Differenzierungsstufe: Unterscheidung nach Identifikation der Messintention der Objektiv-Skalen und entlang der Vertrauensreferenz (a: Messintention identifiziert; b: Messintention verschleiert) bzw. Unterscheidung entlang der Vertrauensreferenz, da in Versuchsgruppe 4 nur die Standard-Skala präsentiert wurde und aufgrund dessen keine Differenzierung nach Identifikation der Messintention vorgenommen werden konnte (c: Vertrauensreferenz entspricht outgroup)

8. Fazit

So sind in *Versuchsgruppe 1* die Messwerte der Objektiv-Skalen bei erfolgreicher Verschleierung der Messintention (Gruppe B) sogar etwas stärker verzerrt ($(.10 \leq r \leq .30)^{93}$ bzw. $(.10 \leq r \leq .70)^{94}$) als bei erfolgreicher Identifikation deren Messintention (Gruppe A) ($(.00 \leq r \leq .50)^{95}$ bzw. $(.00 \leq r \leq .30)^{96}$). Jedoch zeigt ein Blick auf die Korrelationen der Objektiv-Skalen mit den Subkomponenten sozialer Erwünschtheit, dass sie in Gruppe B insgesamt weniger anfällig für sozialerwünschtes Antwortverhalten sind. In Gruppe B ergeben sich keine bis schwache Zusammenhänge mit der SDE- und der IM-Komponente ($.00 \leq r \leq .30$), die sich allerdings nicht wechselseitig kompensieren, sodass insgesamt eine Verfälschung der Messergebnisse resultiert. In Gruppe A ergeben sich dagegen keine bis mittlere Zusammenhänge mit den Subkomponenten ($.00 \leq r \leq .50$), die jedoch gegensätzliche Vorzeichen aufweisen, sodass in Summe eine vergleichsweise geringere Verzerrung der Messwerte resultiert. Bei Kontrolle der Vertrauensreferenz ließ sich dagegen nur für die Objektiv-Kurzskala ein solcher jedoch schwach ausgeprägter Effekt beobachten. (vgl. Tab. 9.1)

In *Versuchsgruppe 2* sind die Messwerte der Objektiv-Skalen bei erfolgreicher Verschleierung der Messintention weniger stark verfälscht ($(.00 \leq r \leq .30)^{97}$ bzw. $(.10 \leq r \leq .30)^{98}$) als bei Identifikation der Messintention ($(.10 \leq r \leq .30)^{99}$ bzw. $(.10 \leq r \leq .30)^{100}$), allerdings scheint dies eine Folge der Kompensation von Selbst- und Fremdtäuschungstendenz zu sein. Für die Objektiv-Kurzskala lässt sich dagegen kein nennenswerter Kompensationseffekt feststellen. Die Verfälschung der Messwerte durch sozialerwünschtes Antwortverhalten bleibt auf dem gleichen Niveau wie bei der Identifikation der Messintention. Bei Kontrolle der Vertrauensreferenz sind nur die Messwerte der Objektiv-Kurzskala in Gruppe B weniger stark verfälscht als in Gruppe A und auch ihre Anfälligkeit für sozialerwünschtes Antwortverhalten hat sich infolge der Intransparenz der Zieldimension reduziert. (vgl. Tab. 9.1)

In *Versuchsgruppe 3* sind die Messwerte der Objektiv-Skalen in Gruppe B ($(.00 \leq r \leq .10)^{101}$ bzw. $(.00 \leq r \leq .10)^{102}$) weniger stark verzerrt als in Gruppe A ($(.10 \leq r \leq .30)^{103}$ bzw. $(.50 \leq r \leq .70)^{104}$). Dieses Resultat entsteht jedoch infolge einer Kompensation von SDE- und IM-Komponente, sodass die Anfälligkeit der Objektiv-Skalen für sozialerwünschtes

⁹³ Wertebereich für Korrelation von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention.

⁹⁴ Wertebereich für Korrelationen von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention und Vertrauensreferenz.

⁹⁵ Wertebereich für Korrelation von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention.

⁹⁶ Wertebereich für Korrelationen von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention und Vertrauensreferenz.

⁹⁷ Wertebereich für Korrelation von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention.

⁹⁸ Wertebereich für Korrelationen von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention und Vertrauensreferenz.

⁹⁹ Wertebereich für Korrelation von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention.

¹⁰⁰ Wertebereich für Korrelationen von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention und Vertrauensreferenz.

¹⁰¹ Wertebereich für Korrelation von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention.

¹⁰² Wertebereich für Korrelationen von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention und Vertrauensreferenz.

¹⁰³ Wertebereich für Korrelation von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention.

¹⁰⁴ Wertebereich für Korrelationen von Objektiv-Skalen und SD-Skala bei Differenzierung nach Identifikation der Messintention und Vertrauensreferenz.

8. Fazit

Antwortverhalten in Gruppe B höher ausfällt als in Gruppe A. Bei Kontrolle der Vertrauensreferenz sind die Messwerte in Gruppe B ebenfalls weniger stark verfälscht als in Gruppe A. Gleichzeitig sinkt die Anfälligkeit der Objektiv-Skala für sozialerwünschtes Antwortverhalten in Gruppe B gegenüber Gruppe A, während die Anfälligkeit der Objektiv-Kurzskala für sozialerwünschtes Antwortverhalten steigt. (vgl. Tab. 9.1)

In Summe sind die Objektiv-Messungen bei erfolgreicher Verschleierung der Messintention weniger stark mit der SD-Skala korreliert, als wenn die Zieldimension der Objektiv-Messungen identifiziert werden konnte. Allerdings geht mit einer geringer ausgeprägten Verfälschung der Messwerte nicht immer auch eine geringer ausgeprägte Anfälligkeit für sozialerwünschtes Antwortverhalten einher. Umgekehrt bedeutet eine geringere Anfälligkeit für sozialerwünschtes Antwortverhalten auch keine zwingende geringere Verzerrung der Messwerte. Insgesamt verbessert die Intransparenz der Zieldimension das Validitätsniveau der Objektiv-Messungen.

Die Messwerte der Standard-Skala sind unabhängig von der Identifikation der Messintention der Objektiv-Skalen vornehmlich dann weniger verfälscht, wenn die Standard-Skala vor den Objektiv-Skalen bearbeitet wurde (*Versuchsgruppe 2*) oder unabhängig von den Objektiv-Skalen bearbeitet wurde (*Versuchsgruppe 4*). Wurde hingegen die Standard-Skala erst nach den Objektiv-Skalen präsentiert (*Versuchsgruppe 1*) sind ihre Messwerte vergleichsweise stärker verfälscht. Es wird daher ein Reihenfolgeeffekt vermutet, nach dem das Antwortverhalten auf der Standard-Skala in *Versuchsgruppe 1* durch die vorangestellten Objektiv-Skalen beeinflusst wird.

So ergeben sich in *Versuchsgruppe 1* in Gruppe A und in Gruppe B jeweils schwache Zusammenhänge zwischen Standard-Skala und SD-Skala ($.10 \leq r \leq .30$). Bei Berücksichtigung der Vertrauensreferenz ergibt sich in Gruppe A ein starker Zusammenhang ($.50 \leq r \leq .70$) und in Gruppe B ein mittlerer Zusammenhang ($.30 \leq r \leq .50$). (vgl. Tab. 9.2)

In *Versuchsgruppe 2* ergibt sich weder in Gruppe A noch in Gruppe B ein nennenswerter Zusammenhang zwischen Standard-Skala und SD-Skala. Auch bei Kontrolle der Vertrauensreferenz bleibt dieses Ergebnis weitestgehend stabil. Allerdings ergibt sich hier in Gruppe A ein schwacher Zusammenhang ($.10 \leq r \leq .30$). (vgl. Tab. 9.2)

Die niedrigen Korrelationen mit der SD-Skala kommen auch im Falle der Standard-Skala häufig infolge von wechselseitiger Kompensation der Subkomponenten sozialer Erwünschtheit zustande. So zeigt sich, dass die Anfälligkeit der Standard-Skala für sozialerwünschtes Antwortverhalten insbesondere bei erfolgreicher Identifikation der Messintention der Objektiv-Skalen, d.h. wenn sowohl die Zieldimension der Objektiv-Skalen als auch der Standard-Skala transparent ist, höher ausfällt, als wenn die Messintention der Objektiv-Skalen erfolgreich verschleiert werden konnte.

Insofern ist anzunehmen, dass zwar die Messwerte der Standard-Messung unabhängig von der Identifikation der Messintention der Objektiv-Messungen eher wenig verfälscht werden. Dieser Effekt scheint jedoch Folge einer Kompensation von SDE- und IM-Komponente zu sein. Gleichzeitig ist die Anfälligkeit der Standard-Skala für sozialerwünschtes Antwortverhalten an die Transparenz der Zieldimension beider Skalentypen gekoppelt. Dieses Ergebnis unterstützt die Annahme eines Reihenfolgeeffektes.

Vergleicht man die Objektiv-Skalen und die Standard-Skala hinsichtlich der Verfälschbarkeit ihrer Messwerte ergibt sich als wesentliches Resultat, dass die erfolgreiche Verschleierung der

8. Fazit

Messintention der Objektiv-Skalen keine Verbesserung des Validitätsniveaus im Kontrast zur Standard-Skala bewirkt. Die Messwerte der Objektiv-Skalen werden nicht weniger durch sozialerwünschtes Antwortverhalten verfälscht als die Messwerte der Standard-Skala. Die Intransparenz der Zieldimension scheint vielmehr dafür zu sorgen, dass die Messergebnisse der verschiedenen Skalentypen ein ähnliches Verzerrungsniveau aufweisen.

So ergeben sich in den *Versuchsgruppen 1* und *2* sowie beim Vergleich der *Versuchsgruppe 3* und *4* bei erfolgreicher Identifikation der Messintention der Objektiv-Skalen Unterschiede in den Zusammenhängen von Objektiv-Skalen und Standard-Skala mit der SD-Skala. Das bedeutet: Wenn die Zieldimensionen beider Skalentypen von den Probanden eingesehen werden konnten, ergaben sich unterschiedliche Verfälschungsniveaus für deren Messwerte. Dabei zeigt sich jedoch auch, dass in *Versuchsgruppe 1* jeweils die Messwerte der Objektiv-Kurzskala am wenigsten verzerrt waren, während in *Versuchsgruppe 2* jeweils die Messwerte der Standard-Skala am wenigsten verzerrt waren. (vgl. Tab. 9.3) Auch beim Vergleich von *Versuchsgruppe 3* und *4* ergaben sich für die Standard-Skala bei Identifikation der Messintention der Objektiv-Skalen vergleichsweise geringere Messwertverzerrungen als für die Objektiv-Skalen. (vgl. Tab. 9.1 und Tab. 9.2) Dies deutet daraufhin, dass die Verzerrungen infolge sozialerwünschten Antwortverhaltens im vorliegenden Fall nicht vollständig unabhängig von der Präsentationsreihenfolge der Vertrauensskalen auftreten. So sind in den *Versuchsgruppen 1* und *2* jeweils die Messwerte der zuerst bearbeiteten Skalentypen weniger verfälscht als die Messwerte der jeweils an zweiter Stelle bearbeiteten Skalentypen. Auch die Anfälligkeit¹⁰⁵ der Vertrauensskalen für sozialerwünschtes Antwortverhalten unterscheidet sich in den verschiedenen Versuchsgruppen. So neigen die Probanden je nach Skala und je nach Versuchsgruppe unterschiedlich stark zu Selbst- und Fremdtäuschung. Hierbei lassen sich jedoch keine Regelmäßigkeiten erkennen.

Bei erfolgreicher Verschleierung der Messintention der Objektiv-Skalen ergaben sich ebenfalls Unterschiede zwischen den Korrelationen der Objektiv-Skalen und der Standard-Skala mit der SD-Skala in und zwischen den verschiedenen Versuchsgruppen. Jedoch fielen diese Unterschiede in den meisten Fällen vergleichsweise marginal aus. Insofern ist bei Intransparenz der Zieldimension der Objektiv-Skalen und Transparenz der Zieldimension der Standard-Skala ein Trend zur wechselseitigen Angleichung des Verzerrungsniveaus der Messwerte der verschiedenen Vertrauensskalen zu erkennen. Dabei sind die Messwerte der Objektiv-Skalen in *Versuchsgruppe 1* weniger stark oder gleich stark verfälscht wie die Messwerte der Standard-Skala, während sie in *Versuchsgruppe 2* stärker oder gleich stark verfälscht sind. (vgl. Tab. 9.3) Beim Vergleich von *Versuchsgruppe 3* und *4* sind je nachdem, ob die Vertrauensreferenz berücksichtigt wird oder nicht, die Messwerte der Standard-Skala oder die Messwerte der Objektiv-Kurzskala in Relation zu den Messwerten der übrigen Skalen weniger stark verzerrt. (vgl. Tab. 9.1 und Tab. 9.2) Auch die Anfälligkeit der Skalentypen für sozialerwünschtes Antwortverhalten unterscheidet sich zwischen den Versuchsgruppen. Dabei überwiegen jedoch die Fälle, in denen die Standard-Skala entweder gleich anfällig wie die Objektiv-Skala oder gleich

¹⁰⁵ Zur Bestimmung der Anfälligkeit der Objektiv-Skalen bzw. der Standard-Skala für sozialerwünschtes Antwortverhalten wurde die Summe des Betrags der Korrelationen der Vertrauensskalen mit den Subkomponenten der BIDR-Kurzskala verwendet. Je höher die Korrelationen mit der Selbst- bzw. Fremdtäuschungskomponente ausfielen, als umso ausgeprägter wurde deren Anfälligkeit für sozialerwünschte Antworten angesehen. Aufgrund wechselseitig entgegengesetzter Vorzeichen der Korrelationen der Vertrauensskalen mit der SDE- und der IM-Komponente ergab sich in vielen Fällen ein Kompensationseffekt der Subkomponenten (Bortz/Döring (2016), S. 439; Gordon/Gross (1978)). In der Folge wurde auf den Vertrauensskalen zwar sehr unaufrichtig geantwortet, jedoch waren die Messwerte in Summe kaum verzerrt (vgl. Korrelation mit gesamter BIDR-Kurzskala).

anfällig wie die Objektiv-Kurzskala für sozialerwünschte Antworten ist bzw. sich annähernd betragsgleiche Zusammenhänge mit den Subkomponenten der BIDR-Kurzskala ergeben.

Hieraus ergibt sich, dass die Messwerte der Objektiv-Messungen und der Standard-Messung nicht gleich stark mit der SD-Skala korreliert sind, wenn die Zieldimension beider Skalentypen transparent ist. Das bedeutet: Auch wenn die Messintention der Objektiv-Skalen identifiziert wurde, sind die Messwerte der Standard-Skala und der Objektiv-Skalen unterschiedlich stark durch sozialerwünschtes Antwortverhalten verzerrt. Dabei scheint die Stärke der Verfälschung mit der Präsentationsreihenfolge der Skalen zusammenzuhängen. So sind jeweils die Messwerte des zuerst bearbeiteten Skalentyps weniger stark verzerrt als die des nachfolgend bearbeiteten Skalentyps. Insofern scheint die Transparenz der Messintention beider Skalentypen einen *Positions-, Reihenfolge-* bzw. *Carry-over*-Effekt nach sich zu ziehen. Auch die Anfälligkeit der Skalentypen für sozialerwünschtes Antwortverhalten unterscheidet sich, allerdings lässt sich hierbei keine nachhaltige Systematik erkennen.

Wurde die Messintention der Objektiv-Messung nicht identifiziert, korrelieren die Objektiv-Skalen und die Standard-Skala mehrheitlich unterschiedlich stark mit der SD-Skala. Dabei kann jedoch nicht festgestellt werden, dass die Messwerte der Objektiv-Skalen systematisch weniger stark durch sozialerwünschtes Antwortverhalten verfälscht werden als die Messwerte der Standard-Skala. Vielmehr ist eine Tendenz zur wechselseitigen Angleichung des Verzerrungsniveaus zu beobachten. Insofern bewirkt die Undurchsichtigkeit der Zieldimension der Objektiv-Skalen keine nachhaltige Verbesserung der Messgüte im Vergleich zur Standard-Skala. Die Objektiv-Skalen und die Standard-Skala scheinen infolge der geringen Augenscheinvalidität der Objektiv-Items eher ähnlich anfällig für sozialerwünschtes Antwortverhalten zu sein. Dabei sind sowohl die Messwerte der Objektiv-Skalen als auch die Messwerte der Standard-Skala vornehmlich dann weniger verfälscht, wenn die Zieldimension der Objektiv-Skalen intransparent bleibt.

8.4.2 Zusammenhang von Messwertdifferenz und sozialer Erwünschtheit

Die Differenz zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala geht zumindest in Teilen auf den Einfluss sozialerwünschten Antwortverhaltens zurück. Dieser Effekt ist unabhängig von der Identifikation bzw. Verschleierung der Messintention. Insofern kann soziale Erwünschtheit als eine Erklärung für den Unterschied zwischen den Messergebnissen dieser beiden Skalentypen herangezogen werden. Für die Differenz zwischen den Messwerten von Objektiv- und Standard-Skala scheint sozialerwünschtes Antwortverhalten dagegen kaum eine bis keine Rolle zu spielen. Für den Unterschied zwischen den Messergebnissen dieser beiden Skalentypen scheinen somit andere Variablen verantwortlich zu sein. Möglicherweise sind hierbei die Unterschiede in der Skalenlänge und in der inhaltlichen Schwerpunktsetzung relevant.

So ergibt sich sowohl in *Versuchsgruppe 1* als auch in *Versuchsgruppe 2* und auch bei Differenzierung nach Identifikation der Messintention sowie der Berücksichtigung der Vertrauensreferenz mit einer Ausnahme kein nennenswerter Zusammenhang der Differenz der Messwerte von Objektiv- und Standard-Skala mit der SD-Skala. Die ermittelten Korrelationen nehmen vornehmlich Werte von $.00 \leq r \leq .10$ an. Nur bei Berücksichtigung der Vertrauensreferenz und unter der Bedingung, dass die Messintention der Objektiv-Skalen nicht

identifiziert wurde, ergibt sich in *Versuchsgruppe 1* ein sehr starker Zusammenhang der genannten Differenz und der SD-Skala ($.70 \leq r \leq 1.00$). Das heißt: Nur in einem einzigen Fall scheint einzig und allein sozialerwünschtes Antwortverhalten für die Messwertunterschiede zwischen Objektiv- und Standard-Skala verantwortlich zu sein. (vgl. Tab. 9.3)

Der Einfluss sozialerwünschten Antwortverhaltens auf die Differenz zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala fällt vergleichsweise größer aus. Über *Versuchsgruppe 1* und 2 hinweg sowie bei Berücksichtigung der Differenzierung nach Identifikation der Messintention sowie der Vertrauensreferenz resultieren hier vornehmlich schwache bis mittlere Zusammenhänge ($.10 \leq r \leq .50$). Eine Ausnahme hiervon ergibt sich wie bereits bei der Objektiv-Skala in *Versuchsgruppe 1*, wenn die Probanden die Messintention der Objektiv-Skalen durchschauen konnten und generalisierte Andere als Bezugsrahmen ihres Vertrauens angegeben hatten. Hier korreliert die Differenz sehr stark mit der SD-Skala ($.70 \leq r \leq 1.00$). (vgl. Tab. 9.3)

Ein weiterer interessanter Effekt ergibt sich entlang der Trennlinie erfolgreicher Identifikation bzw. erfolgreicher Verschleierung der Messintention der Objektiv-Skalen. So scheint die Differenz zwischen den Messwerten der Objektiv-Kurzskala und den Messwerten der Standard-Skala immer dann weniger durch sozialerwünschtes Antwortverhalten hervorgerufen zu werden, wenn die Probanden die Messintention der Objektiv-Skalen nicht identifizieren konnten. Ist dagegen die Messintention beider Skalentypen für die Probanden einsehbar, ergab sich ein vergleichsweise stärkerer Zusammenhang von Differenz und SD-Skala. Bei erfolgreicher Identifikation der Zieldimension der Objektiv-Kurzskala hat sozialerwünschtes Antwortverhalten somit einen größeren Einfluss auf die Messwertunterschiede zwischen den Skalentypen. (vgl. Tab. 9.3)

Die Differenz zwischen den Messwerten der Objektiv-Messungen und den Messwerten der Standard-Messung korrelieren unabhängig von der erfolgreichen Identifikation bzw. Verschleierung der Messintention der Objektiv-Skalen mit der SD-Skala. Die Zusammenhänge sind dabei jedoch maximal mittelstark ausgeprägt und es ergeben sich große Diskrepanzen zwischen Objektiv-Skala und Objektiv-Kurzskala. So scheint nur der Unterschied zwischen den gemittelten Skalensummenwerten der Objektiv-Kurzskala und der Standard-Skala in Teilen auf sozialerwünschtes Antwortverhalten zurückzuführen zu sein. Zusätzlich spielen auch andere Einflüsse eine Rolle. So ist die Differenz zwischen den gemittelten Skalensummenwerten von Objektiv- und Standard-Skala vermutlich vornehmlich auf differenziertere Abfrage der Zieldimension „Generalisiertes Vertrauen“ im Rahmen der Objektiv-Skala im Kontrast zur Standard-Skala zurückzuführen. Somit unterscheiden sich die inhaltliche Schwerpunktsetzung und die Länge dieser beiden Skalen erheblich. Objektiv-Kurzskala und Standard-Skala weisen, wie bereits im Kontext der Validitätsanalyse gezeigt, eine vergleichsweise größere inhaltliche Nähe auf und unterscheiden sich auch in der Skalenlänge bzw. der Ausdifferenzierung weniger stark.

Die Richtung des Zusammenhangs der Differenz zwischen Objektiv- und Standard-Skala mit der SD-Skala scheint unabhängig von der Identifikation der Messintention der Objektiv-Skalen und der Vertrauensreferenz zu sein. So lassen sich zwar vornehmlich positive Zusammenhänge beobachten, diese folgen allerdings keiner nachhaltigen Systematik. Das bedeutet: Sozialerwünschtes Antwortverhalten bewirkt hier überwiegend eine Vergrößerung der Diskrepanz zwischen den Messwerten von Objektiv- und Standard-Skala. Wie aus dem vorangegangenen Abschnitt ersichtlich, fallen die Korrelationswerte von Differenz und SD-Skala

8. Fazit

im Falle der Objektiv-Skala jedoch so klein aus, dass an dieser Stelle von keinem nennenswerten Einfluss sozialer Erwünschtheit ausgegangen werden muss. Insofern ist auch die Richtung des Zusammenhangs vernachlässigbar, da soziale Erwünschtheit offensichtlich nicht die zentrale Erklärungsvariable für die Messwertunterschiede zwischen diesen beiden Skalentypen darstellt. (vgl. Tab. 9.3)

Hinsichtlich der Differenz zwischen Objektiv-Kurzskala und Standard-Skala lässt sich eine entlang der Trennlinie erfolgreiche Identifikation bzw. Verschleierung der Messintention der Objektiv-Skalen verlaufende Systematik erkennen. So ergeben sich in *Versuchsgruppe 1* zwar unabhängig von der Differenzierung der Fälle nach Identifikation der Messintention und Vertrauensreferenz ausschließlich positive Zusammenhänge von Differenz und SD-Skala. Das bedeutet: Der Unterschied zwischen den Messwerten der beiden Skalentypen vergrößert sich grundsätzlich infolge sozialerwünschten Antwortverhaltens. Es kann jedoch festgestellt werden, dass dieser Effekt bei erfolgreicher Verschleierung der Messintention gemindert wird. So fallen die Korrelationswerte zwischen Differenz und SD-Skala bei Intransparenz der Zieldimension der Objektiv-Skalen niedriger aus, als wenn deren Zieldimension von den Probanden identifiziert werden konnte. Bei Wirksamkeit der für die Objektiv-Items angestrebten geringen Augenscheininvalidität bezüglich der Zieldimension „Generalisiertes Vertrauen“ verringert sich der Effekt sozialerwünschten Antwortverhaltens auf die Differenz zwischen den Messwerten der Skalentypen bzw. sozialerwünschte Antworten tragen weniger stark zur Vergrößerung der Messwertunterschiede bei. In *Versuchsgruppe 2* ergeben sich bei Transparenz der Messintention von Objektiv-Kurzskala und Standard-Skala positive Zusammenhänge zwischen der Messwertdifferenz von Objektiv-Kurzskala und Standard-Skala und der SD-Skala. Bei erfolgreicher Verschleierung der Messintention der Objektiv-Skalen ergeben sich dagegen negative Zusammenhänge zwischen Differenz und SD-Skala. Auch hier fallen die Korrelationswerte bei Intransparenz der Zieldimension der Objektiv-Skalen tendenziell niedriger aus, als wenn deren Zieldimension von den Probanden identifiziert werden konnte. Bei Transparenz der Zieldimensionen beider Skalentypen vergrößern sich somit die Messwertunterschiede zwischen Objektiv-Kurzskala und Standard-Skala infolge sozialerwünschten Antwortverhaltens. Können die Probanden dagegen die Messintention der Objektiv-Kurzskala nicht aufdecken, so verringern sich die Messwertunterschiede zwischen Objektiv-Kurzskala und Standard-Skala infolge sozialerwünschten Antwortverhaltens, während gleichzeitig der Einfluss sozialer Erwünschtheit auf die Differenz zwischen den beiden Skalentypen zurückgeht. (vgl. Tab. 9.3)

Außerdem fällt auf, dass bei erfolgreicher Verschleierung der Messintention die Korrelationen der Differenz von Objektiv-Kurzskala und Standard-Skala mit der SD-Skala in *Versuchsgruppe 1* jeweils ein positives Vorzeichen aufweisen, während sie in *Versuchsgruppe 2* ein negatives Vorzeichen besitzen. (vgl. Tab. 9.3) Es kann daher nicht ausgeschlossen werden, dass die Richtung des Effektes sozialerwünschten Antwortverhaltens in diesem Fall von der Präsentationsreihenfolge der Skalentypen unabhängig ist. Möglicherweise liefert die Standard-Skala, die in *Versuchsgruppe 2* vor der Objektiv-Kurzskala bearbeitet wurde, einen Hinweis auf die Zieldimension der Objektiv-Kurzskala, sodass eine bewusste und zielgerichtete Angleichung des Antwortverhaltens auf der Objektiv-Kurzskala an das Antwortverhalten auf der Standard-Skala stattfindet. Ein umgekehrter Reihenfolgeeffekt in *Versuchsgruppe 1* ist dagegen weniger wahrscheinlich, da die Standard-Skala dort erst an zweiter Stelle bearbeitet und die Zieldimension der Objektiv-Kurzskala erfolgreich verschleiert wurde. Insofern konnte sich aus der Bearbeitung der Objektiv-Kurzskala kein Hinweis auf die gemeinsame Zieldimension der

beiden Skalentypen ergeben. Dagegen spricht außerdem, dass sich in *Versuchsgruppe 1* die Messwertdifferenz infolge sozialerwünschten Antwortverhaltens tendenziell vergrößert und die Probanden somit eher zu einem voneinander verschiedenen Antwortverhalten auf den unterschiedlichen Vertrauensskalen zu neigen scheinen.

Insgesamt ist die Differenz zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala über die Gruppengrenzen und alle Differenzierungsstufen hinweg mehrheitlich stärker durch sozialerwünschtes Antwortverhalten beeinflusst als die Differenz zwischen den Messwerten von Objektiv- und Standard-Skala. (vgl. Tab. 9.3) Dieser Befund kann als weiterer Hinweis darauf gewertet werden, dass soziale Erwünschtheit nicht als vordergründige Ursache für die Messwertabweichungen zwischen Objektiv- und Standard-Skala herangezogen werden kann, sondern möglicherweise inhaltliche Disparitäten oder die Unterschiede in der Ausdifferenzierung der Abfrage der Zieldimension „Generalisiertes Vertrauen“ für die Messwertdifferenzen verantwortlich sind. Hinsichtlich der Objektiv-Kurzskala hat bereits die Validitätsanalyse gezeigt, dass sie eine vergleichsweise größere Nähe zur Standard-Skala aufweist als die Objektiv-Skala. Entsprechend spielen zur Erklärung der Messwertdifferenzen weniger potentielle inhaltliche Abweichung als Variablen wie soziale Erwünschtheit bzw. reaktive Effekte im Allgemeinen eine Rolle. So leistet sozialerwünschtes Antwortverhalten keinesfalls eine vollständige Aufklärung der Messwertunterschiede, sondern es müssen weitere Einflussfaktoren in Betracht gezogen werden, die in dieser Untersuchung aber nicht beachtet werden konnten bzw. wurden.

Unter der Bedingung, dass die Messintention der Objektiv-Messungen identifiziert wurde, korreliert die Differenz zwischen den Messwerten der Objektiv-Messungen und der Standard-Messung vornehmlich positiv mit der SD-Skala. Es besteht insofern bei den Probanden eine Neigung zu inkonsistentem Antwortverhalten auf den verschiedenen Skalentypen, wenn deren Zieldimension einsehbar ist. Es findet somit überwiegend ein Auseinanderdriften der Messwerte von Objektiv-Skalen und Standard-Skala infolge sozialerwünschten Antwortens statt. Dabei ist der Effekt sozialer Erwünschtheit hinsichtlich der Differenz zwischen Objektiv- und Standard-Skala jedoch zu vernachlässigen.

Unter der Bedingung, dass die Messintention der Objektiv-Messungen nicht identifiziert wurde, korreliert die Differenz zwischen den Messwerten der Objektiv-Messungen und der Standard-Messungen gleichermaßen häufig positiv wie negativ mit der SD-Skala. Insofern kann bei Intransparenz der Zieldimension der Objektiv-Skalen keine eindeutige Systematik im Antwortverhalten der Probanden festgestellt werden. Sie neigen weder zu übermäßig konsistentem noch zu übermäßig inkonsistentem Antwortverhalten auf den verschiedenen Skalentypen, wenn die Messintention der Objektiv-Skalen erfolgreich verschleiert werden konnte. Dabei scheint sich soziale Erwünschtheit nur auf die Differenz zwischen Objektiv-Kurzskala und Standard-Standard-Skala nachhaltig auszuwirken, während der Effekt für die Differenz zwischen Objektiv- und Standard-Skala vergleichsweise gering ausfällt. Hinsichtlich des Zusammenhangs der Differenz der Messwerte von Objektiv-Kurzskala und Standard-Skala mit der SD-Skala scheint allerdings ein versuchsgruppenspezifischer Effekt zu existieren, der auf die Präsentationsreihenfolge der Skalentypen zurückzuführen ist. Vermutet wird, dass die Standard-Skala einen Hinweis auf die gemeinsame Messintention der verschiedenen Skalentypen gibt und sich sozialerwünschtes Antwortverhalten infolgedessen in Versuchsgruppe 2 in Form einer Angleichung derer Messwerte manifestiert.

8. Fazit

Hinsichtlich der Entstehung des Unterschieds zwischen den Messwerten der Objektiv-Skala und der Standard-Skala infolge sozialerwünschten Antwortverhaltens lässt sich keine belastbare Systematik ableiten. In der Tendenz scheint jedoch vornehmlich das Antwortverhalten auf der Objektiv-Skala oder eine Wechselwirkung des Antwortverhaltens auf beiden Skalentypen für die Messwertdifferenz verantwortlich zu sein. Dagegen kann nur in einem einzigen Fall der Messwertunterschied auf das Antwortverhalten auf der Standard-Skala zurückgeführt werden. In diesem Zusammenhang ist außerdem zu bedenken, dass sozialerwünschtes Antwortverhalten nur einen verschwindend geringen Effekt auf die Messwertdifferenz zwischen Objektiv- und Standard-Skala hat. Die beobachteten Messwertunterschiede sind somit wohl überwiegend auf von sozialer Erwünschtheit verschiedene Einflüsse zurückzuführen. (vgl. Tab. 9.3)

Über die *Versuchsgruppen 1* und *2* hinweg betrachtet, treten unabhängig von der Identifikation der Messintention der Objektiv-Skalen, der Berücksichtigung der Vertrauensreferenz und der Tatsache, dass soziale Erwünschtheit keinen nennenswerten Effekt auf die Messwertdifferenz hat, die folgenden Konstellationen auf: In einigen Fällen weist die Objektiv-Skala eine stärkere Korrelation mit der SD-Skala auf als die Standard-Skala, wobei die Richtung der jeweiligen Zusammenhänge deckungsgleich sind. Die Messwerte der Skalen vergrößern sich dabei simultan durch sozialerwünschte Antworten. (vgl. Tab. 9.3)

Aus der Analyse der Mittelwertvergleiche wissen wir, dass die Messwerte der Objektiv-Skala grundsätzlich niedriger ausfallen als die Messwerte der Standard-Skala. In der Folge weist die Differenz zwischen den Skalen eine negative Korrelation mit der SD-Skala auf. Das bedeutet: Der Abstand zwischen den Messwerten der Skalentypen verringert sich aufgrund sozialerwünschten Antwortverhaltens. Da die Objektiv-Skala stärker durch sozialerwünschtes Antwortverhalten beeinflusst wird als die Standard-Skala und ihre Messwerte deshalb stärker nach oben korrigiert werden, scheint vornehmlich das Antwortverhalten auf der Objektiv-Skala für diesen Effekt verantwortlich zu sein. (vgl. Tab. 9.3)

Fällt die Korrelation der Standard-Skala mit der SD-Skala größer als die der Objektiv-Skala mit der SD-Skala aus und weisen beide Zusammenhänge positive Vorzeichen auf oder korreliert die Objektiv-Skala stärker mit der SD-Skala als die Standard-Skala und weisen beide Zusammenhänge negative Vorzeichen auf, so resultiert für den Zusammenhang von Messwertdifferenz und SD-Skala ein positives Vorzeichen. Auch hierbei spielt die Tatsache, dass die Messwerte der Objektiv-Skala grundsätzlich niedriger ausfallen als die der Standard-Skala, eine Rolle. Im ersten Fall werden die Messwerte der Standard-Skala durch den Einfluss sozialerwünschten Antwortverhaltens stärker gesteigert als die der Objektiv-Skala, sodass die Mittelwertdifferenz gesteigert wird. Im zweiten Fall werden die Messwerte der Objektiv-Skala infolge sozialerwünschter Antworten stärker verringert als die Messwerte der Standard-Skala, sodass auch hier eine Vergrößerung des Unterschieds aufgrund sozialerwünschten Antwortverhaltens resultiert. (vgl. Tab. 9.3)

Daneben existiert ein Fall, bei dem die Standard-Skala nicht mit der SD-Skala korreliert und die Objektiv-Skala einen negativen Zusammenhang mit der SD-Skala aufweist. Auch hier ist die positive Korrelation der Messwertdifferenz mit der SD-Skala, d.h. das Auseinanderdriften der Messwerte als Folge sozialerwünschten Antwortverhaltens, auf die Objektiv-Skala zurückzuführen, da sich deren Messwerte unter dem Einfluss sozialer Erwünschtheit verringern. (vgl. Tab. 9.3)

Weisen die Zusammenhänge der Einzelskalen wechselseitig entgegengesetzte Vorzeichen auf. Das heißt: Steigen die Messwerte der einen Skala durch den Einfluss sozialerwünschten Antwortverhaltens, während die der anderen sinken, so korreliert die Messwertdifferenz

8. Fazit

negativ mit der SD-Skala. Der Messwertunterschied reduziert sich somit infolge sozialer Erwünschtheit.

Fallen die Korrelationen der Einzelskalen mit der SD-Skala annähernd gleich groß aus, entscheiden womöglich schon leichte Unterschiede in der Stärke der Zusammenhänge darüber, ob sich die Mittelwertdifferenz infolge sozialerwünschten Antwortverhaltens vergrößert oder verringert.

Die Abweichungen zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala infolge sozialerwünschten Antwortverhaltens scheinen, unabhängig von der Identifikation bzw. Verschleierung der Messintention und unabhängig von der Vertrauensreferenz, vornehmlich durch das Antwortverhalten auf der Standard-Skala motiviert zu sein. Dabei wird auf der Standard-Skala jeweils das wahre Vertrauen überschätzt, während auf der Objektiv-Kurzskala vergleichsweise aufrichtiger geantwortet wird. Die Objektiv-Kurzskala leistet insofern im Kontrast zur Standard-Skala einen Beitrag zur Verbesserung der Messgüte. Bei dieser Interpretation ist zu berücksichtigen, dass sich im Rahmen der Mittelwertvergleiche gezeigt hat, dass die Messwerte der Objektiv-Kurzskala grundsätzlich niedriger ausfallen als die Messwerte der Standard-Skala. (vgl. Tab. 9.3)

So zeigt sich innerhalb von *Versuchsgruppe 1*, dass die Standard-Skala entweder stärker oder annähernd gleich stark mit der SD-Skala korreliert ist wie die Objektiv-Kurzskala. Die Einzelskalen weisen dabei ebenso wie die zugehörige Messwertdifferenz einen positiven Zusammenhang mit der SD-Skala auf. Das bedeutet: Mit steigender Tendenz zu sozialerwünschtem Antwortverhalten steigert sich das auf den Einzelskalen angegebene Vertrauensniveau bzw. die Probanden neigen auf beiden Skalentypen zu einer Überschätzung ihres wahren Vertrauens, wenn sie sozialerwünscht antworten. Auch der Messwertunterschied zwischen der Objektiv-Kurzskala und der Standard-Skala vergrößert sich infolge sozialerwünschten Antwortverhaltens. Da die Standard-Skala stärker durch soziale Erwünschtheit beeinflusst wird als die Objektiv-Kurzskala, scheint die Differenz zwischen deren Messwerten somit überwiegend dem Antwortverhalten auf der Standard-Skala anzulasten zu sein. (vgl. Tab. 9.3)

In *Versuchsgruppe 2* ist die Objektiv-Kurzskala entweder stärker oder gleich stark mit der SD-Skala korreliert wie die Standard-Skala. Dabei weisen Standard- und SD-Skala in einem Fall keinen und ansonsten einen negativen Zusammenhang auf. Auf der Standard-Skala unterschätzen die Probanden somit ihr wahres Vertrauen tendenziell infolge sozialerwünschten Antwortens. Zwischen der SD-Skala und der Objektiv-Skala ergeben sich sowohl negative als auch positive Zusammenhänge. Dabei lässt sich jedoch keine entlang der Identifikation der Messintention und der Vertrauensreferenz verlaufende Systematik erkennen. Insofern lässt sich auch für die Verfälschung des per Objektiv-Kurzskala ermittelten Vertrauensniveaus keine eindeutige Tendenz ableiten. Interessanterweise weist der Zusammenhang der Messwertdifferenz der genannten Skalen mit der SD-Skala mehrheitlich das gleiche Vorzeichen wie der Zusammenhang von Standard- und SD-Skala auf. Zusätzlich tritt dieser Befund immer dann auf, wenn das Vorzeichen des Zusammenhangs von Objektiv-Kurzskala und SD-Skala dem Vorzeichen des Zusammenhangs von Standard- und SD-Skala diametral entgegengesetzt ist. Das bedeutet: Der Messwertunterschied zwischen Objektiv-Kurzskala und Standard-Skala verringert sich immer dann unter dem Einfluss sozialerwünschten Antwortverhaltens, wenn die Probanden ihr wahres Vertrauensniveau auf der Objektiv-Kurzskala überschätzen bzw. ihr wahres Vertrauensniveau auf der Standard-Skala unterschätzen. Auf den ersten Blick scheint vornehmlich das Antwortverhalten auf der Objektiv-Kurzskala für diese Angleichung der Messwerte der beiden Skalentypen verantwortlich zu sein. Vor dem Hintergrund des im

vorherigen Abschnitt thematisierten potentiellen Reihenfolgeeffektes in *Versuchsgruppe 2* ist jedoch eher zu vermuten, dass dieser Befund Ausdruck des Versuchs der Probanden ist, ihre Antworten auf der Objektiv-Kurzskala an ihre Antworten auf der Standard-Skala anzupassen. Insofern ist mit hoher Wahrscheinlichkeit auch in *Versuchsgruppe 2* eher das Antwortverhalten auf der Standard-Skala für die Messwertdifferenz von Objektiv-Kurzskala und Standard-Skala verantwortlich als das Antwortverhalten auf der Objektiv-Kurzskala. (vgl. Tab. 9.3)

Unabhängig davon, ob die Messintention der Objektiv-Skala erfolgreich identifiziert oder verschleiert werden konnte, scheinen in der Tendenz Richtung und Stärke des Einflusses sozialer Erwünschtheit auf die Differenz zwischen den Messwerten der Objektiv- und der Standard-Skala vornehmlich auf die Beeinflussung der Messwerte der Objektiv-Skala durch sozialerwünschtes Antwortverhalten zurückzuführen zu sein. Insofern trägt die Objektiv-Skala im Vergleich mit der Standard-Skala nicht zur Steigerung der Validität der Messergebnisse bei.

Unabhängig davon, ob die Messintention der Objektiv-Kurzskala erfolgreich identifiziert oder verschleiert werden konnte, scheinen in der Tendenz Richtung und Stärke des Einflusses sozialer Erwünschtheit auf die Differenz zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala vornehmlich auf die Beeinflussung der Messwerte der Standard-Skala durch sozialerwünschtes Antwortverhalten zurückzuführen zu sein. Insofern leistet die Objektiv-Kurzskala im Vergleich mit der Standard-Skala prinzipiell einen Beitrag zur Verbesserung der Messgüte.

8. Fazit

Tab. 8.3: Einfluss sozialerwünschten Antwortverhaltens auf die Einzelskalen und die Messwertdifferenzen zwischen Objektiv- und Standard-Messung

VG	Diff.-stufe	Relation von $ r_{S,BIDR} $, $ r_{OK,BIDR} $ und $ r_{OKS,BIDR} $	Wertebereich von $ r_{S,BIDR} $, $ r_{OK,BIDR} $ und $ r_{OKS,BIDR} $	Relation von $\sum r_{S,SDE;IM} $, $\sum r_{OK,SDE;IM} $ und $\sum r_{OKS,SDE;IM} $	Relation von $ r_{Diff(OK/S)} $ und $ r_{Diff(OKS/S)} $	Wertebereich von $ r_{Diff(OK/S)} $ und $ r_{Diff(OKS/S)} $
1	1	$OKS(+) < S(+) < OK(+)$	$.10 \leq r \leq .50$	$OKS < S < OK$	$Diff(OK/S)(+) < Diff(OKS/S)(+)$	$.00 \leq r \leq .30$
	2a	$OKS(+) < S(+) < OK(+)$	$.00 \leq r \leq .50$	$S < OK < OKS$	$Diff(OK/S)(-) < Diff(OKS/S)(+)$	$.00 \leq r \leq .50$
	2b	$OK(+) \approx OKS(+) \approx S(+)$	$.10 \leq r \leq .30$	$OKS < S < OK$	$Diff(OK/S)(+) < Diff(OKS/S)(+)$	$.00 \leq r \leq .30$
	3a	$OKS(+) < OK(+) < S(+)$	$.00 \leq r \leq .70$	$OK < S < OKS$	$Diff(OK/S)(+) > Diff(OKS/S)(+)$	$.70 \leq r \leq 1.00$
	3b	$OKS(+) < S(+) < OK(+)$	$.10 \leq r \leq .70$	$S < OK = OKS$	$Diff(OK/S)(-) < Diff(OKS/S)(+)$	$.00 \leq r \leq .30$
2	1	$S(-) \approx OK(-) \approx OKS(+)$	$.00 \leq r \leq .10$	$OK < S < OKS$	$Diff(OK/S)(+) \approx Diff(OKS/S)(-)$	$.00 \leq r \leq .10$
	2a	$S(0) < OK(-) < OKS(-)$	$.00 \leq r \leq .30$	$OK < S < OKS$	$Diff(OK/S)(+) < Diff(OKS/S)(+)$	$.00 \leq r \leq .30$
	2b	$S(-) \approx OK(+) < OKS(+)$	$.00 \leq r \leq .30$	$S < OKS < OK$	$Diff(OK/S)(-) < Diff(OKS/S)(-)$	$.00 \leq r \leq .30$
	3a	$S(-) \approx OK(-) < OKS(-)$	$.10 \leq r \leq .70$	$S < OK < OKS$	$Diff(OK/S)(-) < Diff(OKS/S)(+)$	$.00 \leq r \leq .50$
	3b	$S(-) < OK(-) \approx OKS(+)$	$.00 \leq r \leq .30$	$S \approx OKS < OK$	$Diff(OK/S)(+) < Diff(OKS/S)(-)$	$.00 \leq r \leq .30$

VG: Versuchsgruppe

 $|r_{S,BIDR}|$ bzw. (S): Betrag der Korrelation der Standard-Skala mit der SD-Skala (hier: BIDR-Kurzskala) $|r_{OK,BIDR}|$ bzw. (OK): Betrag Korrelation der Objektiv-Skala mit der SD-Skala (hier: BIDR-Kurzskala) $|r_{OKS,BIDR}|$ bzw. (OKS): Betrag Korrelation der Objektiv-Kurzskala mit der SD-Skala (hier: BIDR-Kurzskala) $\sum |r_{S,SDE;IM}|$: Summe des Betrags der Korrelationen der Standard-Skala mit der Selbst- (SDE) und Fremdtäuschungskomponente (IM) der BIDR-Kurzskala $\sum |r_{OK,SDE;IM}|$: Summe des Betrags der Korrelationen der Objektiv-Skala mit der Selbst- (SDE) und Fremdtäuschungskomponente (IM) der BIDR-Kurzskala $\sum |r_{OKS,SDE;IM}|$: Summe des Betrags der Korrelationen der Objektiv-Kurzskala mit der Selbst- (SDE) und Fremdtäuschungskomponente (IM) der BIDR-Kurzskala $|r_{Diff(OK/S)}|$: Betrag der Korrelation der Messwertdifferenz von Objektiv- und Standard-Skala mit der SD-Skala $|r_{Diff(OKS/S)}|$: Betrag der Korrelation der Messwertdifferenz von Objektiv-Kurzskala und Standard-Skala mit der SD-Skala

(+: positiver Zusammenhang)

(-): negativer Zusammenhang

1. Differenzierungsstufe: Alle gültigen Fälle

2. Differenzierungsstufe: Unterscheidung nach Identifikation der Messintention der Objektiv-Skalen (a: Messintention identifiziert; b: Messintention verschleiert)

3. Differenzierungsstufe: Unterscheidung nach Identifikation der Messintention der Objektiv-Skalen und entlang der Vertrauensreferenz (a: Messintention identifiziert; b: Messintention verschleiert)

8.5 Radiusproblem

Beim Radiusproblem handelt es sich um die zentrale Problemstellung der Vertrauensmessung. Das ursprüngliche Radiusproblem befasste sich dabei mit der Frage, ob die Phrase „*most people*“, die im Kontext der Items der *Trust in People-Scale* (hier: Standard-Skala) verwendet wird, von allen Probanden in der gleichen Art und Weise interpretiert wird. So fanden sich in empirischen Untersuchungen Evidenzen dafür, dass manche Probanden spezifische Andere bzw. die *ingroup* als Referenz ihres Vertrauens annahmen, während andere Probanden generalisierte Andere bzw. die *outgroup* als Bezugsrahmen ihres Vertrauens nannten. Dies hatte zur Folge, dass mit dem gleichen Messinstrument zwei verschiedene Formen des Vertrauens erfasst wurden, die nicht ohne Weiteres voneinander differenziert werden konnten, sodass das resultierende Vertrauensniveau nicht eindeutig Partikularem oder Generalisiertem Vertrauen zugeordnet werden konnte.

Das Radiusproblem wurde später auch noch auf die Phrase „*most of the time*“ erweitert und beinhaltete somit neben einer referentiellen Dimension zusätzlich noch eine zeitliche Dimension. Diese zeitliche Dimension des Radiusproblems setzt bei der Fraglosigkeit und Universalität Generalisierten Vertrauens an. So wurde angemerkt, dass Vertrauen abhängig vom jeweiligen Vertrauensobjekt ist. Bei der Art und Weise der Formulierung der Standarditems bleibt aber, wie der Bezugsrahmen des Vertrauens, auch der Gegenstand des Vertrauens unbestimmt. Aus den Antworten der Probanden lässt sich nicht bestimmen, welches Vertrauensobjekt Basis für das angegebene Vertrauen ist oder ob sich das Vertrauen nur auf spezifische Vertrauensobjekte bezieht.

In der Summe ergibt sich aus dem Radiusproblem somit eine gewichtige Bedrohung der Validität von Vertrauensmessungen. So war es auch Ziel der vorliegenden Untersuchung zu überprüfen, inwiefern die spezielle Item-Architektur der Objektiv-Skalen einen Beitrag zur Lösung des Radiusproblems leisten kann. Hierbei wurde sich allerdings auf die ursprüngliche Version des Radiusproblems beschränkt und geprüft, ob Probanden bei der Bearbeitung der Objektiv-Skalen eher zu generalisierten Anderen als Vertrauensreferenz neigen als zu spezifischen Anderen. Außerdem wurde untersucht, ob der Anteil derjenigen, die die *outgroup* als Bezugspunkt ihres Vertrauens nennen, bei der Messung von Vertrauen mit Hilfe der Objektiv-Skalen größer ausfällt als bei der Vertrauensmessung per Standard-Skala.

Der Effekt des Radiusproblems wurde sowohl in der Pilotstudie als auch in der Hauptstudie analysiert. Für die Pilotstudie gilt dabei jedoch die bereits im Rahmen der Validitätsbestimmung beschriebene Problematik. So beziehen sich die Angaben der Probanden bezüglich ihrer Vertrauensreferenz hier strenggenommen auf die 42 Items umfassende Rohskala und nicht auf die 14 bzw. 7 Items umfassenden resultierenden Objektiv-Skalen. Zudem wurde innerhalb der Pilotstudie kein Vergleich mit dem Standardinstrument der Vertrauensmessung realisiert.

Die Analyseergebnisse der Pilotstudie zeigen, dass zwar die Mehrzahl der Probanden bei der Bearbeitung der Objektiv-Skalen generalisierte Andere als Bezugsrahmen ihres Vertrauens genutzt hatten, sich die Anteile der Vertrauensreferenzen *outgroup* (59,4%) und *ingroup* (40,6%) jedoch kaum unterscheiden. Dies wird bei Berechnung der 99%-Konfidenzintervalle für die jeweiligen Anteilswerte deutlich. So schwankt der wahre Wert des *ingroup*- und des *outgroup*-Anteils jeweils um 8,91% um die empirisch ermittelten Anteilswerte. Hieraus ergibt sich zwar keine Überschneidung des Wertebereichs beider Anteile, allerdings kommen sich die Wertebereiche sehr nahe.

Es wird dementsprechend davon ausgegangen, dass bei der Messung von Vertrauen mit Hilfe der Objektiv-Skalen die Anteile von *ingroup* und *outgroup* als Vertrauensreferenz annähernd gleich verteilt sind bzw. nicht in der Mehrzahl der Fälle generalisierte Andere als Bezugsrahmen des Vertrauens angegeben werden.

Im Rahmen der Hauptstudie konnte sowohl der Bezugsrahmen des per Objektiv-Skalen gemessenen Vertrauens als auch der Bezugsrahmen des per Standard-Skala gemessenen Vertrauens bestimmt werden. Dabei zeigte sich, dass in jeder Versuchsgruppe und unabhängig vom Skalentyp vornehmlich generalisierte Andere als Bezugspunkt des Vertrauens angegeben wurden. Dies stellt einen weiteren Hinweis darauf dar, dass sowohl die Objektiv-Skalen als auch die Standard-Skala vornehmlich Generalisiertes Vertrauen abbilden. Bei den Objektiv-Skalen schwankte der Anteil der *ingroup* zwischen 28% und 40,8%, während der Anteil der *outgroup* Werte zwischen 59,2% und 72% annahm. Bei der Standard-Skala gaben zwischen 32% und 42,9% der Probanden die *ingroup* als Vertrauensreferenz an, während zwischen 57,1% und 68% der Probanden die *outgroup* als Bezugsrahmen nannten. Die Auswertung der 99%-Konfidenzintervalle für die Anteilswerte zeigte jedoch, dass zwischen den wahren Anteilswerten von *in-* und *outgroup* kaum bzw. kein Unterschied besteht. So überschneiden sich die Wertebereiche der Konfidenzintervalle für die wahren Anteile von *in-* und *outgroup* in *Versuchsgruppe 1* sowohl für die Objektiv- als auch für die Standard-Skala. In *Versuchsgruppe 2* lassen sich zwar weder für die Objektiv-Skalen noch für die Standard-Skala eine Überschneidung der Intervalle der Anteilswerte beobachten, allerdings fallen die Konfidenzintervalle sehr breit aus und die untere Grenze des *outgroup*-Anteils liegt für beide Skalentypen verhältnismäßig nahe an der oberen Grenze des *ingroup*-Anteils. Es ist somit mit einer großen Streubreite für die wahren Anteilswerte zu rechnen. In *Versuchsgruppe 3* wurden ausschließlich die Objektiv-Skalen bearbeitet. Hier ergab sich eine Überlappung der Wertebereiche der Anteile von *in-* und *outgroup*, aus der sogar ein umgekehrtes Anteilsverhältnis für die wahren Anteilswerte resultieren könnte. Das bedeutet: In Wahrheit könnten die Objektiv-Skalen möglicherweise sogar eher Partikulares als Generalisiertes Vertrauen abbilden. In *Versuchsgruppe 4* wurde den Probanden ausschließlich die Standard-Skala präsentiert und auch hier ergab sich bei der Auswertung der 99%-Konfidenzintervalle eine Überschneidung der Wertebereich für die wahren Anteilswerte von *in-* und *outgroup* als Vertrauensreferenz. Vergleicht man die Ergebnisse aus den *Versuchsgruppe 3* und *4* zeigt sich, dass wie bereits in den *Versuchsgruppen 1* und *2* kein wesentlicher Unterschied zwischen den Objektiv-Skalen und der Standard-Skala besteht, was die Verteilung der Anteile der Vertrauensreferenzen angeht. Bei beiden Skalentypen werden in einem ähnlichen Verhältnis zueinander spezifische und generalisierte Andere als Bezugspunkt des Vertrauens genannt.

Sowohl bei der Messung von Vertrauen anhand der Objektiv-Skalen als auch bei der Messung von Vertrauen mit Hilfe der Standard-Skala werden mehrheitlich generalisierte Andere als Vertrauensreferenz angegeben. Innerhalb der *Versuchsgruppen 1* und *2* und bei Vergleich der *Versuchsgruppen 3* und *4* ergeben sich dabei kaum Unterschiede hinsichtlich der Anteilswerte von *in-* und *outgroup* zwischen den Skalentypen. Insofern ist anzunehmen, dass bei der Messung von Vertrauen mit Hilfe der Objektiv-Skalen in etwa gleich häufig generalisierte Andere als Vertrauensreferenz angegeben werden wie bei der Vertrauensmessung per Standard-Skala. Die Verwendung der Objektiv-Skalen im Rahmen der Vertrauensmessung trägt somit nicht zur Lösung des Radiusproblems bei. Vielmehr scheint sich das Radiusproblem bei den Objektiv-Skalen und der Standard-Skala in gleicher Art und Weise zu manifestieren.

8.6 Schlussfolgerung und Ausblick

Die Designvorgaben der Hybridklasse konnten bei der Erstellung der Objektiv-Skalen weitestgehend erfolgreich umgesetzt werden. So erfolgte die Datenerhebung im Rahmen einer standardisierten

8. Fazit

Situation und anhand eines standardisierten Untersuchungsinstrumentes, das einem Fragebogen entsprach.

Die Analyseergebnisse aus Pilot- und Hauptstudie zeigten zudem, dass die Probanden die Zieldimension der Objektiv-Skalen mehrheitlich nicht durchschauen konnten. Insofern konnte auch die Vorgabe, Items zu verwenden, die keine mit der Messintention übereinstimmende Augenscheinvalidität aufweisen, weitestgehend umgesetzt werden. Es ist jedoch zu bemerken, dass die Undurchschaubarkeit der Zieldimension nicht vollständig gewährleistet werden konnte. Es sollte daher eine Revision der Itemauswahl stattfinden und gegebenenfalls Item-Neuformulierungen, -eliminationen oder -substitutionen vorgenommen werden, sodass in Zukunft eine vollständige Verschleierung der Messintention erreicht wird.

Die Abfrage der Zieldimension „Generalisiertes Vertrauen“ entsprach insofern keiner Selbstbeurteilung, als dass die Probanden hauptsächlich um Angaben hinsichtlich Dritter bzw. Menschen im Allgemeinen oder die Evaluation von Situationen, in die sie selbst nicht unmittelbar involviert waren bzw. die einen allgemeinen Charakter aufwiesen, gebeten wurden und keine explizite Auskunft über die Person der Probanden betreffende Aspekte erforderlich war. Allerdings erfolgte die Abfrage in Form einer Ratingskala, sodass die Probanden dazu angehalten waren, selbst anzugeben, wie sie die Aussagen der einzelnen Items bewerten. Das bedeutet: Der Anspruch, keine Selbstbeurteilung durch die Probanden zuzulassen, konnte ebenfalls nur teilweise erfüllt werden. Diesem Umstand ließe sich nur mit einer radikalen Umstellung des Messkonzeptes entgegenwirken. Ziel dieser Studie war es allerdings, ein Messverfahren zu entwickeln, das dem Fragebogenformat entspricht und sich gleichzeitig weniger anfällig für reaktive Effekte zeigt. Grund hierfür war, dass es sich bei standardisierten Interviews um die meistgenutzte Erhebungsform innerhalb der empirischen Sozialforschung handelt und sich entsprechende Studiendesigns verglichen mit anderen Erhebungsverfahren leicht konzeptionieren und ausführen lassen. Zusätzlich wird die Generierung einer großen Fallzahl ermöglicht, was sich positiv auf die Validität der Ergebnisse auswirkt. Eine Abkehr vom Fragebogenformat würde somit dem Forschungsziel widersprechen und die mit dem Fragebogenformat verbundenen Vorteile zunichtemachen.

Sowohl die Objektiv-Skalen als auch die Standard-Skala stellen reliable Messverfahren dar. Die Messgenauigkeit der Skalen liegt dabei auf einem ähnlich hohen Niveau. Allerdings scheinen die Objektiv-Skalen eine differenzierte Erfassung der Zieldimension im Vergleich zur Standard-Skala zu ermöglichen, die aber auch mit einer leichten Reduktion der Messpräzision im Kontrast zur Standard-Skala einhergeht. Von den Objektiv-Skalen scheint die kompaktere Objektiv-Kurzskala eine präzisere Abbildung Generalisierten Vertrauens zu erlauben.

Für die Objektiv-Skalen konnte Inhalts- Kriteriums- und Konstruktvalidität bezüglich des anvisierten Zielkonstruktes festgestellt werden. Sie ermöglichen insofern in gleicher Weise wie die Standard-Skala (*Trust in People-Scale*) eine valide Abbildung der Zieldimension „Generalisiertes Vertrauen“. Beide Objektiv-Skalen weisen eine ähnliche Faktorstruktur auf wie die Standard-Skala, die in letzter Instanz in einer einfaktoriellen Lösung konvergiert. Vor dem Hintergrund dieses Ergebnisses wird vermutet, dass es sich bei den durch die *Trust in People-Scale* abgefragten Dimensionen (*Vertrauensbereitschaft*, *Fairness* und *Hilfsbereitschaft*) um die wesentlichen Kernaspekte Generalisierten Vertrauens handelt. Aus der Analyse der Konstruktvalidität ließ sich außerdem ableiten, dass zwischen Objektiv-Kurzskala und Standard-Skala ein stärkerer Zusammenhang bzw. eine größere inhaltliche Nähe besteht als zwischen Objektiv- und Standard-Skala. Entsprechend der Definition konvergenter Validität sollte die Objektiv-Kurzskala somit eine im Vergleich zur Objektiv-Skala bessere Abbildung Generalisierten Vertrauens liefern.

Bei erfolgreicher Verschleierung der Messintention sind die Messwerte der Objektiv-Skalen tendenziell weniger stark durch sozialerwünschtes Antwortverhalten verfälscht, als wenn die Probanden die Messintention der Objektiv-Skalen identifizieren konnten. Dieser Effekt ist allerdings für die Objektiv-Skala deutlich stärker ausgeprägt als für die Objektiv-Kurzskala. Gleichzeitig wurde jedoch festgestellt, dass die Messwerte der Objektiv-Skalen nicht systematisch weniger stark durch sozialerwünschtes Antwortverhalten verzerrt werden als die Messwerte der Standard-Skala. Insofern scheint die Intransparenz der Zieldimension durchaus einen kleinen positiven Effekt auf die Messgüte der Objektiv-Skalen zu haben, gleichzeitig lässt sich allerdings keine Verbesserung der Validität der Messergebnisse im Kontrast zur Standard-Skala feststellen. Vielmehr scheinen die Objektiv-Messungen und die Standard-Messung in Summe unabhängig von der Identifikation der Messintention der Objektiv-Skalen eher gleich stark durch soziale Erwünschtheit beeinflusst zu sein. So sind die Objektiv-Skalen und die Standard-Skala in den meisten Fällen ähnlich stark mit der SD-Skala korreliert oder die Standard-Skala ist stärker als die Objektiv-Kurzskala, aber schwächer als die Objektiv-Skala mit der SD-Skala korreliert (vgl. Tab. 9.3).

Ein weiterer interessanter Befund ergab sich bei Betrachtung der Korrelationen der Objektiv-Skalen und der Standard-Skala mit den Subkomponenten sozialer Erwünschtheit, Selbsttäuschung (SDE) und Fremdtäuschung (IM). Es zeigte sich, dass bei den Objektiv-Skalen jeweils in 60% der Fälle und bei der Standard-Skala in 50% der Fälle eine Reduktion der Messwertverfälschung infolge einer wechselseitigen Kompensation der SDE- und der IM-Komponente stattfand. Selbst- und Fremdtäuschungstendenz der Probanden wirkten hierbei einander entgegen, sodass in der Summe aufrichtiger geantwortet wurde. Teilweise wurden dadurch die Verzerrungen der Messwerte infolge sozialerwünschten Antwortverhaltens auf nahezu null reduziert. Insbesondere bei der Objektiv-Kurzskala war dieser Effekt mit der erfolgreichen Verschleierung der Messintention verknüpft. Dagegen trat bei der Standard-Skala der Effekt häufiger bei Transparenz der Zieldimension der Objektiv-Skalen auf. Für die Objektiv-Skala ließ sich keine nachhaltige Systematik entdecken. Die Probanden scheinen somit auf der Objektiv-Kurzskala aufgrund der Intransparenz der Zieldimension unbewusst aufrichtiger zu antworten. Es wird daher vermutet, dass es den Probanden tendenziell weniger möglich ist, zielgerichtete Verfälschungen ihres Antwortverhaltens herbeizuführen, wenn die „demand characteristics“ der Messung nicht klar definiert sind. Mit anderen Worten: Die Anwendung insbesondere der Objektiv-Kurzskala kann zwar nicht verhindern, dass die Probanden sozialerwünscht antworten, allerdings scheinen bewusste Verzerrungen infolge sozialerwünschten Antwortverhaltens reduziert werden zu können. Durch die geringe Augenscheinvalidität der Items der Objektiv-Kurzskala wird dabei möglicherweise eine klare Definition der Zieldimension verhindert, was ein systematisches sozialerwünschtes Antwortverhalten erschwert.

Die Resultate der Mittelwertvergleiche im Rahmen des Survey-Experimentes zeigen, dass sich zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala signifikante Unterschiede ergeben. Dabei liegt das von den Objektiv-Skalen ermittelte Vertrauensniveau jeweils unterhalb des von der Standard-Skala ermittelten Vertrauensniveaus. Gleichzeitig scheinen die überzufälligen Differenzen zwischen den Messwerten der verschiedenen Skalentypen an die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen gekoppelt zu sein. Bei erfolgreicher Identifikation der Messintention der Objektiv-Skalen ergeben sich dagegen kaum überzufällige Unterschiede zwischen den Messwerten der Objektiv-Skalen und der Standard-Skala. Das bedeutet: Wenn die Probanden sowohl die Zieldimension der Objektiv-Skalen als auch die Zieldimension der Standard-Skala einsehen können, stellen die verschiedenen Skalentypen vornehmlich gleichwertige Messverfahren dar bzw. es ergeben sich annähernd äquivalente Vertrauensniveaus. Die erfolgreiche Verschleierung der Messintention der Objektiv-Skalen bewirkt dagegen, dass sich die Probanden auf

den Objektiv-Skalen weniger vertrauensvoll darstellen als auf der Standard-Skala. Es wird daher vermutet, dass das auf den Objektiv-Skalen angegebene Vertrauen eher dem wahren Vertrauensniveau der Probanden entspricht als das Vertrauen, welches sie auf der Standard-Skala angegebenen haben. Zu bemerken ist allerdings, dass der Unterschied zwar überzufällig ist, in absoluten Zahlen ausgedrückt aber verhältnismäßig gering ausfällt. So erreicht die Differenz zwischen den gemittelten Skalensummenwerten maximal .75 Skaleneinheiten.

Weitere Analyseschritte zeigen, dass die Differenz zwischen den Messwerten der Skalentypen unabhängig von der Identifikation der Messintention zumindest im Falle der Objektiv-Kurzskala teilweise auf den Einfluss sozialerwünschten Antwortverhaltens zurückgeführt werden kann. Auf die Messwertdifferenz zwischen Objektiv- und Standard-Skala hat sozialerwünschtes Antwortverhalten dagegen keinen nachhaltigen Effekt. Hier resultieren die wechselseitigen Messwertabweichungen voraussichtlich aus den unterschiedlichen inhaltlichen Differenzierungsniveaus der beiden Skalentypen und den voneinander verschiedenen Skalensummen. Insofern scheint die überzufällige Differenz zwischen den gemittelten Skalensummenwerten der Objektiv-Kurzskala und der Standard-Skala Ausdruck von Validitätsunterschieden der jeweils resultierenden Messwerte zu sein. Wie bereits beschrieben wird vermutet, dass die Messwerte der Objektiv-Kurzskala eher dem wahren Vertrauensniveau entsprechen als die Messwerte der Standard-Skala.

Setzt man den Zusammenhang von Messwertdifferenz und sozialer Erwünschtheit mit dem Effekt sozialerwünschten Antwortverhaltens auf die Einzelskalen in Beziehung zeigt sich, dass unabhängig von der Identifikation der Messintention die Differenz zwischen den Messwerten der Objektiv-Kurzskala und der Standard-Skala vornehmlich auf die Beeinflussung der Messwerte der Standard-Skala durch sozialerwünschtes Antwortverhalten zurückzuführen ist. Wobei die Probanden ihr Vertrauensniveau auf der Standard-Skala tendenziell überschätzen, während sie auf der Objektiv-Kurzskala vergleichsweise aufrichtiger antworten. Lässt man außer Acht, dass die Messwertdifferenz zwischen Objektiv- und Standard-Skala kaum mit der SD-Skala zusammenhängt, zeigt sich, dass hier vornehmlich das Antwortverhalten auf der Objektiv-Skala für die Messwertunterschiede verantwortlich ist. Somit trägt nur die Objektiv-Kurzskala gegenüber der Standard-Skala zu einer Verbesserung der Messgüte bei. Dabei ist jedoch zu beachten, dass sowohl die Messwerte der Objektiv-Kurzskala als auch die Messwerte der Standard-Skala durch sozialerwünschtes Antwortverhalten verfälscht werden. Auch scheinen die speziellen Designvorgaben für die Objektiv-Kurzskala und hierbei im Speziellen die geringe Augenscheinvalidität der Items bezüglich der Zieldimension keinen Effekt zu haben.

Die Auswahl des theoretischen Konstruktes „Generalisiertes Vertrauen“ als Zieldimension für die im Rahmen dieser empirischen Untersuchung vorgenommene Integration und Kontrastierung direkter und indirekter Messverfahren bzw. eines Hybriden aus beiden Messverfahrensgruppen stellte sich im Nachhinein als problematisch dar und es kann nicht vollständig ausgeschlossen werden, dass sich allein aus der Verwendung dieser Zieldimension Messwertunterschiede zwischen den Erhebungsmethoden oder Messwertverzerrungen im Allgemeinen ergeben haben. Verantwortlich hierfür war das sogenannte Radiusproblem der Vertrauensforschung. So wird nur dann tatsächlich Generalisiertes Vertrauen gemessen, wenn die Probanden generalisierte Andere als Bezugsrahmen ihres Vertrauens angeben. Werden dagegen spezifische Andere als Vertrauensreferenz genannt, erfasst das eingesetzte Messinstrument Partikulares Vertrauen. In der Konsequenz kann bei der einfachen Abfrage des Vertrauensniveaus mit Hilfe des Standardinstruments, der *Trust in People-Scale*, nicht bestimmt werden, ob Partikulares oder Generalisiertes Vertrauen gemessen wird. Dieser Umstand wiegt deshalb so schwer, weil das Partikulare Vertrauen in der Regel höher ausfällt als das Generalisierte Vertrauen. Mit anderen Worten: Menschen bringen Personen aus ihrem unmittelbaren sozialen Umfeld, die ihnen

8. Fazit

persönlich bekannt sind, ein größeres Vertrauen entgegen, als Personen, denen sie erst ein einziges Mal begegnet sind oder Fremden. Die *Trust in People-Scale* bildet somit nicht exklusiv Generalisiertes Vertrauen ab, sondern es findet eine Diffusion zweier voneinander zu unterscheidender Spielarten des Vertrauens statt, in deren Folge eine Verzerrung der resultierenden Messwerte geschehen kann. Das wahre Niveau Generalisierten Vertrauens wird dann möglicherweise verfälscht und tendenziell überschätzt.

Sowohl im Kontext der Querschnittsbefragung der Pilotstudie als auch im Kontext des Survey-Experimentes der Hauptstudie wurde das Radiusproblem kontrolliert. In beiden Fällen zeigte sich, dass die Objektiv-Skalen keinen nachhaltigen Beitrag zur Lösung des Radiusproblems leisten konnten. So gab zwar die Mehrzahl der Probanden generalisierte Andere als Referenzgruppe ihres Vertrauens an, d.h. in den meisten Fällen bildeten die Objektiv-Skalen tatsächlich das Niveau des individuellen Generalisierten Vertrauens ab, jedoch fiel der Anteil derjenigen, die spezifische Andere als Vertrauensreferenz nannten, nicht wesentlich geringer aus. Weitere Analysen zeigten, dass sich die Konfidenzintervalle für die wahren Anteilswerte überschneiden.

Im Rahmen des Survey-Experimentes konnte gezeigt werden, dass sich das Vertrauensniveau der Probanden, die spezifische Andere als Vertrauensreferenz angegeben hatten, nicht vom Vertrauensniveau derer unterscheidet, die generalisierte Andere als Vertrauensreferenz genannt hatten. Insofern sind keine Messwertverzerrungen infolge der Vermischung der Vertrauensdimensionen zu erwarten. Trotzdem wurde die undifferenzierte Abfrage von Partikularem und Generalisiertem Vertrauen als potentielle Störvariable für die Analyse der Mittelwertvergleiche und die vergleichende Analyse des Einflusses sozialerwünschten Antwortverhaltens identifiziert und entsprechend beachtet. Strenggenommen dürften für ein abschließendes Urteil über die gefundenen Messwertunterschiede zwischen Objektiv-Skalen und Standard-Skala sowie den Effekt sozialerwünschten Antwortverhaltens auf diese Differenz somit nur die Ergebnisse zu Rate gezogen werden, die sich bei Kontrolle der Vertrauensreferenz eingestellt haben. Die Beschränkung der in die Analysen eingehenden Fälle auf Probanden, die ausschließlich generalisierte Andere als Bezugsrahmen ihres Vertrauens genannt hatten, hatte jedoch eine starke Reduktion der Stichprobengröße zur Folge. Entsprechend sind die Ergebnisse auf dieser Differenzierungsstufe weniger belastbar, als wenn alle gültigen Fälle in die Analyse eingehen würden. Für eine endgültige Absicherung der Ergebnisse könnte eine Replikationsstudie, für die ein anderes theoretisches Konstrukt herangezogen wird, sorgen.

Die Forschungsergebnisse zeigen zudem, dass das im Rahmen der Objektiv-Skalen verwendete Konzept der indirekten Abfrage einer interessierenden Zieldimension vornehmlich auf latente Variablen angewendet werden sollte und bei der Erfassung manifester Variablen an ihre Grenzen gerät bzw. weniger gut funktioniert. Insofern beinhalten die Objektiv-Skalen zwar die Möglichkeit der Adaption auf unterschiedlichste Inhalte und die Option der spezifischen Abfrage einer Zieldimension, bleiben aber trotzdem auf bestimmte theoretische Konstrukte oder empirische Phänomene limitiert. Der Grund hierfür liegt darin, dass es sich bei latenten Variablen um nicht direkt beobachtbare Merkmale handelt, die in der Regel anhand von Indikatoren erfasst werden. Die Ausprägungen eines latenten Merkmals werden dabei aus beobachtbaren Sachverhalten bzw. manifesten Variablen abgeleitet, entsprechend erfolgt die Ermittlung des Wertes einer latenten Variable prinzipiell auf indirektem Wege. Im Rahmen der Objektiv-Skalen erfolgt die Operationalisierung der Zieldimension mittels Korrelaten der Messintention entlang von mit der Messintention verwandten Aspekten oder unter Zuhilfenahme von mit der Messintention assoziiertem Verhalten, d.h. anhand von Indikatoren der Zieldimension. Allerdings wird dabei eine explizite Benennung des interessierenden Konstruktes bzw. direkte Hinweise auf die Messintention vermieden. Damit entspricht das Erhebungskonzept der Objektiv-Skalen im Wesentlichen dem von latenten Variablen.

Dagegen handelt es sich bei manifesten Variablen um direkt beobachtbare Merkmale, die einer unmittelbaren Abfrage zugänglich sind. Manifeste Variablen stellen in diesem Sinne oft ihre eigene Messvorschrift dar und werden nicht anhand von stellvertretenden Charakteristika erfasst. Ein indirekter Zugang zu manifesten Variablen ist vergleichsweise schwer zu bewerkstelligen bzw. macht eine erhebliche Menge an Zusatzinformationen notwendig, enthält eine hohe Gefahr für Fehlschlüsse, ist fast zwangsläufig mit einem Informationsverlust gegenüber einer direkten Abfrage verbunden und stellt vor allem keine besonders ökonomische Vorgehensweise dar. Möchte man beispielsweise Informationen darüber erhalten, welche finanziellen Mittel einer Person monatlich zur Verfügung stehen, so wird normalerweise entweder eine offene Abfrage des monatlichen Einkommens vorgenommen oder das Einkommensniveau wird anhand der Einordnung in Einkommenskategorien geschätzt, um reaktive Effekte einzudämmen. Eine indirekte Abfrage des Einkommensniveaus könnte dagegen beispielsweise entlang der materiellen Güter erfolgen, die eine Person besitzt. Verfügt die Person über vergleichsweise teure Gegenstände oder leistet sich regelmäßig kostenintensive Aktivitäten oder Produkte, kann dies als Hinweis auf ein hohes Einkommen gewertet werden. Es ist jedoch leicht ersichtlich, dass es sich hierbei um keine sonderlich belastbare Evidenz für die Verortung einer Person in einer bestimmten Einkommenskategorie handelt. So könnte der materielle Besitz einer Person auch von einer Erbschaft oder Schenkung herrühren und ist möglicherweise weniger stark mit dem individuellen Einkommen verknüpft. Um die Person einer Einkommensklasse zuordnen zu können, müssten zusätzlich zunächst Informationen dahingehend eingeholt werden, welchen Wert die von der Person angegebenen materiellen Güter besitzen. Wertvolle materielle Güter können zusätzlich verschiedenen Klassen (z.B. Aktien, Immobilien, Automobile, Edelsteine- oder -metalle) angehören. Um eine disjunkte und exhaustive Abfrage zu gewährleisten, muss dann entweder eine Vielzahl von Items verwendet werden oder das verwendete Item eine große Anzahl von Ausprägungen aufweisen. Im Vergleich zu einer direkten Abfrage des Einkommens, die mit einem einzelnen Item auskommt, handelt es sich bei einer indirekten Abfrage des Einkommensniveaus somit um eine vergleichsweise ineffiziente Erhebungsvariante. Am Schluss kann eine indirekte Abfrage des Einkommens lediglich den Informationsgehalt einer ordinalskalierten Variable erreichen, da sich aus den Informationen über den materiellen Besitz einer Person allenfalls Aussagen über das Einkommensniveau, nicht aber über das exakte Einkommen einer Person ableiten lassen. Dagegen ermöglicht eine direkte Abfrage die Erfassung des monatlichen Einkommens auch auf metrischem Niveau bzw. die Angabe eines exakten numerischen Wertes, was einem im Vergleich höheren Informationsgehalt entspricht. Diese Aufschlüsselung unvermeidlicher Vorarbeiten und Begleiterscheinungen der indirekten Abfrage einer manifesten Variable lässt sich auf beliebige direkt beobachtbare Merkmale übertragen und unterscheidet sich dabei höchstwahrscheinlich nur im zu betreibenden Aufwand bei der Formulierung einer angemessenen Messvorschrift.

Wie bereits beschrieben, stellt einer der großen Vorteile direkter Messverfahren die vergleichsweise schnelle Konzeptionierung entsprechender Untersuchungsinstrumente dar. Dieser Vorzug konnte bei der hier vorgenommenen Integration von direkten und indirekten Messverfahren nur unzureichend bewahrt werden. Ein großes Hindernis stellte hierbei die Notwendigkeit einer detaillierten theoretischen Aufarbeitung des interessierenden Konstruktes dar, die die Grundlage für die Formulierung, der innerhalb der Objektiv-Skalen verwendeten subtilen Items darstellte. Um die Benutzerfreundlichkeit des Erhebungskonzeptes der Objektiv-Skalen für Folgeuntersuchungen oder Adaptionen auf andere Zieldimensionen zu verbessern, wird daher vorgeschlagen, sich in Zukunft stärker an bereits bestehenden Messinstrumenten zu orientieren. Für viele theoretische Konstrukte existieren oft bereits entsprechende Standardinstrumente, die deren wesentlichen und relevanten

8. Fazit

Aspekte abfragen. Hieraus ergeben sich prinzipiell zwei Strategien zur Übersetzung dieser direkten Messverfahren in eine Objektiv-Skala:

- (1) Die bestehenden Items könnten so umformuliert werden, dass sie keinen Hinweis mehr auf die Messintention der Skala geben.
- (2) Entlang der vom Standardinstrument abgefragten Dimensionen könnten neue Items formuliert werden, die keine mit der Zieldimension übereinstimmende Augenscheinvalidität aufweisen.

In jedem Fall sollte der resultierenden Skala eine Coverstory vorangestellt werden, um die Intransparenz des Messprinzips zu unterstützen. Zwar entbindet diese Art des Vorgehens bei der Erstellung einer Objektiv-Skala nicht von der Anfertigung eines entsprechenden theoretischen Überbaus, jedoch wird die Anzahl der zu berücksichtigenden Quellen drastisch reduziert. So basieren die Standardinstrumente oft auf den bezüglich des interessierenden Merkmals am häufigsten rezipierten Denkschulen bzw. es wurde ein einzelnes klar definiertes theoretisches Konzept als Basis des Messinstruments herangezogen oder ein spezifischer Forscher hat das Messinstrument als Konsequenz seiner Forschungstätigkeit erstellt.

Ein vorteilhafter Nebeneffekt ist außerdem, dass sich die Reliabilität und die Validität der resultierenden Objektiv-Skalen erhöhen dürften. Dieser Effekt würde dabei vornehmlich auf die inhaltliche Angleichung von Standardinstrument und Objektiv-Skala zurückgehen.

Zum Schluss lässt sich konstatieren, dass das im Rahmen dieser Studie vorgestellte Verfahren zur Integration direkter und indirekter Messverfahren zu diesem Zeitpunkt noch als zu unausgereift erscheint, als dass es als eine Alternative zu den bereits bestehenden Erhebungsmethoden verwendet werden sollte. Gleichwohl haben bereits die Ergebnisse der vorliegenden Untersuchung Ansätze dahingehend erkennen lassen, dass das Zielkonstrukt „Generalisiertes Vertrauen“ von den Objektiv-Skalen ähnlich präzise und valide erfasst wird, wie von der Standard-Skala (*Trust in People-Scale*). Die Objektiv-Skalen und die Standard-Skala stellen somit tendenziell mindestens gleichwertige Messverfahren dar. Ein besonderer Vorteil der Objektiv-Skalen liegt jedoch darin, dass sie eine interessierende Zieldimension valide abfragen können, ohne dass die Messintention offengelegt werden muss. Dagegen macht eine valide Messung mit Hilfe der Standard-Skala eine Transparenz der Messintention notwendig. Insofern zeigen die Objektiv-Skalen, dass den Probanden nicht unbedingt bewusst sein muss, welche Zieldimension mit Hilfe eines Messinstruments abgefragt wird, um gültige Messergebnisse zu erhalten.

Trotzdem wurde das Ziel, mit Hilfe der Objektiv-Skalen eine Reduktion von Messwertverfälschungen infolge reaktiver Effekte zu erreichen und damit eine Verbesserung der Messgüte gegenüber direkten Messverfahren zu erzielen, verfehlt. Zwar zeigten die Analyseergebnisse, dass insbesondere die Messergebnisse der Objektiv-Kurzskala durchaus weniger verfälscht und dadurch valider sein können als die Messergebnisse der Standard-Skala, allerdings existieren noch eine Reihe intervenierender Variablen und Störgrößen, die innerhalb dieser Studie nicht berücksichtigt werden konnten oder noch zu wenig Beachtung erfahren haben, oder teilweise auch erst im Lauf der Studie zu Tage traten. So zeigte sich, dass sozialerwünschtes Antwortverhalten allenfalls einen Bruchteil der Messwertdifferenz zwischen den Skalentypen aufklären kann. Folgeuntersuchungen sollten sich dieser Probleme annehmen und auch die in diesem Abschnitt vorgeschlagenen Verbesserungsmöglichkeiten beachten, um den vielversprechenden Ansatz eines Hybriden aus direkten und indirekten Messverfahren, der tatsächlich die Vorzüge beider Verfahrensgruppen vereint, während er ihre Nachteile ausgleicht, weiterzuentwickeln.

Dabei ergeben sich gleich eine Reihe von weiterführenden Forschungsansätzen. So könnte untersucht werden, welche weitere intervenierenden Variablen neben sozialerwünschtem Antwortverhalten für die Differenz zwischen den Objektiv-Skalen und der Standard-Skala verantwortlich sind. Es könnte

8. Fazit

untersucht werden, ob sich die hier für Generalisiertes Vertrauen gefundenen Resultate für andere Konstrukte replizieren lassen. Zudem könnte analysiert werden, inwiefern sich Modeeffekte bei der Anwendung des hier vorgestellten Designs Objektiver Tests einstellen. Mit anderen Worten ergeben sich Unterschiede zwischen einer Umsetzung als PAPI und Selbstausfüller zu Erhebungsmodi bei denen ein Interviewer anwesend ist (z.B. CAPI, CATI) oder spielt möglicherweise das Medium der Befragung eine Rolle (z.B. CAWI). Zum Schluss könnte auch der Effekt Objektiver Tests für andere Arten systematischer Verzerrungen aufgearbeitet werden. Möglicherweise können neben sozialerwünschtem Antwortverhalten auch andere Bias eingedämmt oder sogar verhindert werden.

Literaturverzeichnis

- Abdi, H.; Edelman, B.; Valentin, D.; Dowling, W. J. (2009): *Experimental Design and Analysis for Psychology*. Oxford: Oxford University Press
- Albrecht, G. (1972): Zur Stellung historischer Forschungsmethoden und nicht-reaktiver Methoden im System der empirischen Sozialforschung. *KZfSS, Sonderheft 16: Soziologie und Sozialgeschichte. Aspekte und Probleme*, S. 242-293
- Alesina, A.; La Ferrara, E. (2000): The Determinants of Trust. NBER Working Paper No. 7621
- Alesina, A.; La Ferrara, E. (2002): Who trusts others? *Journal of Public Economics*, 85, S. 207-234
- Amelang, M.; Gold, A.; Külbel, E. (1984): Über einige Erfahrungen mit einer deutschsprachigen Skala zur Erfassung zwischenmenschlichen Vertrauens (Interpersonal Trust). *Diagnostica*, 30(3), S. 198-215
- Anderson, J.R. (1983): *The Architecture of Cognition*. Cambridge: Harvard University Press
- Arendasy, M.; Sommer, M.; Herle, M.; Schützhofer, B.; Inwanschitz, D. (2011): Modeling Effects of Faking on an Objective Personality Test. *Journal of Individual Differences*, 32(4), S. 210-218
- Asendorpf, J.B.; Banse, R.; Mücke, D. (2002): Double Dissociation between Implicit and Explicit Personality Self-Concept: The Case of Shy Behavior. *Journal of Personality and Social Psychology*, 83(2), S. 380-393
- Axhausen, S. (1999): Projektive Verfahren. In: Roth, E.; Heidenreich, K.; Holling, H. (Hrsg.): *Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis*. München: Oldenbourg
- Babbie, E. R. (1975): *The Practice of Social Research*. Belmont: Wadsworth Publishing Company
- Bacharach, M.; Gambetta, D. (2001): Trust in Signs. In: Cook, K. (Hrsg.): *Trust in Society*. New York: Russell Sage Foundation
- Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R. (2006): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Heidelberg: Springer
- Baier, A. (1986): Trust and Antitrust. *Ethics*, 96, S. 231-260
- Baier, A. (1992): Trusting People. *Philosophical Perspectives*, 6, S. 137-153
- Barbalet, J.M. (1993): Confidence: Time and Emotion in the Sociology of Action. *Journal for the Theory of Social Behaviour*, 23(3), S. 229-247
- Barbalet, J.M. (1996): Social Emotions: Confidence, Trust and Loyalty. *International Journal of Sociology and Social Policy*, 16(9/10), S. 75-96
- Barbalet, J.M. (1998): Emotion, Social Theory, and Social Structure. A macrosociological approach. Cambridge: Cambridge Univ. Press
- Barbalet, J.M. (2009): A characterization of trust, and its consequences. In: *Theory and Society*, 38(4), S. 367-382
- Barber, B. (1983): *The logic and limits of trust*. New Brunswick, N.J.: Rutgers University Press

- Beauducel, A. (2001): Problems with parallel analysis in data sets with oblique simple structure. *Methods of Psychological Research*, 6(2), S. 141-157
- Beierlein, C.; Kemper, C.J.; Kovaleva, A.; Rammstedt, B. (2012): Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen (KUSIV3). GESIS-Working Papers 2012/22
- Berg, B.L. (2009): *Qualitative Research Methods for the Social Sciences*. Boston: Allyn and Bacon
- Bernard, H. R. (2013): *Social Research Methods. Qualitative and Quantitative Approaches*. Thousand Oaks: Sage Publications
- Bierhoff, H.-W. (2002): Politisches Vertrauen: Verschiedene Dimensionen, verschiedene Ebenen der Betrachtung. In: Schmalz-Bruns, R.; Zintl, R. (Hrsg.): *Politisches Vertrauen. Soziale Grundlagen reflexiver Kooperation*. Baden-Baden: Nomos
- Bjørnskov, C. (2003): The Happy Few: Cross-Country Evidence on Social Capital and Life Satisfaction. *Kyklos*, 56(1), S. 3-16
- Bjørnskov, C. (2006): Determinants of generalized trust: A cross-country comparison. *Public Choice*, 130, S. 1-21
- Boeckmann, R.; Tyler, T. (2002): Trust, Respect, and the Psychology of Political Engagement. *Journal of Applied Social Psychology*, 32(10), S. 2067-2088
- Bornstein, R.F.; Rossner, S.C.; Hill, E.L.; Stepanian, M.L. (1994): Face Validity and Fakability of Objective and Projective Measures of Dependency. *Journal of Personality Assessment*, 63(1), S. 363-386
- Bortz, J. (2005): *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer
- Bortz, J.; Döring, N. (2006): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin: Springer
- Bortz, J.; Döring, N. (2016): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin: Springer
- Bortz, J.; Lienert, G. A.; Boehnke, K. (2008): *Verteilungsfreie Methoden in der Biostatistik*. Heidelberg: Springer
- ~~Bourdieu, P. (1983): Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In: Kreckel, R. (Hrsg.): Soziale Ungleichheiten, Göttingen: Schwartz~~
- Bradburn, N. M. (1988): Response Effects. In: Rossi, P. H.; Wright, J. D.; Anderson, A. B. (Hrsg.): *Handbook of Survey Research*. Orlando: Academic Press
- Braver, M. C.; Braver, S.L. (1988): Statistical Treatment of the Solomon Four-Group Design: A Meta-Analytic Approach. *Psychological Bulletin*, 104(1), S. 150-154
- Brehm, J.; Rahn, W. (1997): Individual-Level Evidence for Causes and Consequences of Social Capital. *American Journal of Political Science*, 41(3), S. 999-1023
- Brewer, M. B. (1981): Ethnocentrism and Its Role in Interpersonal Trust. In: Brewer, M.B.; Collins, B.E. (Hrsg.): *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey-Bass

- Brewer, M. B. (2008): Depersonalized Trust and Ingroup Cooperation. In: Krueger, J. (Hrsg.): *Rationality and Social Responsibility. Essays in Honor of Robert Mason Dawes*. New York: Psychology Press
- Briggs, S. R.; Cheek, J., M. (1986): The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54(1), S. 106-148
- Brockner, J.; Siegel, P.; Daly, J.; Tyler, T.; Martin, C. (1997): When Trust Matters: The Moderating Effect of Outcome Favorability. *Administrative Science Quarterly*, 42, S. 558-583
- Bröder, A. (2011): *Versuchsplanung und experimentelles Praktikum*. Göttingen: Hogrefe
- Brosius, F. (2011): *SPSS 19*. Heidelberg: mitp
- Brosius, H.-B.; Koschel, F. Haas, A. (2012): *Methoden der empirischen Kommunikationsforschung. Eine Einführung*. Wiesbaden: VS-Verlag
- Brunel, F.F.; Tietje, B.C.; Greenwald, A.G. (2004): Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition? *Journal of Consumer Psychology*, 14(4), S. 385-404
- Bryant, F. B.; Yarnold, P. R. (1994): Principal-Components Analysis and Exploratory and Confirmatory Factor Analysis. In: Grimm, L. G.; Yarnold, P. R. (Hrsg.): *Reading and Understanding Multivariate Statistics*. Washington: American Psychological Association
- Bühner, M. (2006): *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium
- Bühner, M. (2011): *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium
- Bühner, M.; Ziegler, M. (2009): *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson
- Bungard, W. (1984): *Sozialpsychologische Forschung im Labor*. Göttingen: Hogrefe
- Bungard, W.; Lück, H.E. (1974): *Forschungsartefakte und nicht-reaktive Messverfahren*. Stuttgart: Teubner
- Bungard, W.; Lück, H.E. (1995): Nichtreaktive Verfahren. In: Flick, U.; von Kardorff, E.; Keupp, H.; von Rosenstiel, L.; Wolff, S. (Hrsg.): *Handbuch Qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen*. Weinheim: Beltz
- Butler, J.K. (1991): Toward Understanding and Measuring Conditions of Trust: Evaluation of a Conditions of Trust Inventory. *Journal of Management*, 17(3), S. 643-663
- Campbell, D.T. (1957): Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), S. 297-312
- Campbell, D.T.; Stanley, J. C. (1969): *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally
- Cattell, R.B. (1941): An objective test of character-temperament I. *Journal of General Psychology*, 25, S. 59-73
- Cattell, R.B. (1944): An objective test of character-temperament I. *Journal of Social Psychology*, 19, S. 99-114

- Cattell, R.B. (1946): Description and measurement of personality. New York: World Book
- Cattell, R.B. (1955): Handbook for the objective-analytic personality test batteries: (including adult and child O-A batteries). Savoy: Institute for Personality and Ability Testing
- Cattell, R.B. (1958): What is „Objective“ in „Objective Personality Tests“? *Journal of Counseling Psychology*, 5(4), S. 285-289
- Cattell, R.B. (1966): The Scree Test for the Number of Factors. *Multivariate Behavioral Research*, 1, S. 245-276
- Cattell, R.B. (1978): The scientific use of factor analysis. New York: Plenum
- Cattell, R.B.; Warburton, F.W. (1967): Objective Personality and Motivation Tests. A Theoretical Introduction and Practical Compendium. Urbana: University of Illinois Press
- Charness, G.; Gneezy, U.; Kuhn, M.A. (2012): Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81, S. 1-8
- Chaudhuri, A. (2011): Randomized Response and Indirect Questioning Techniques in Surveys. Boca Raton: Chapman&Hall/CRC
- Chaudhuri, A.; Christofides, T.C. (2013): Indirect Questioning in Sample Surveys. Berlin: Springer
- Christie, R.; Geis, F. (1970): Studies in Machiavellianism. New York: Academic Press Inc.
- Chun, K.; Campbell, J. (1974): Dimensionality of the Rotter Interpersonal Trust Scale. *Psychological Reports*, 35, S. 1059-1070
- Chun, K.; Campbell, J. (1975): Notes on the Internal Structure of Wrightsman's Measure of Trustworthiness. *Psychological Reports*, 37, S. 323-330
- Clark, H.H.; Schober, M.F. (1992): Asking Questions and Influencing Answers. In: Tanur, J. M. (Hrsg.): Questions about Questions. Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation
- Cohen, J. (1988): Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Erlbaum
- Cohen, J. (1992): A Power Primer. *Psychological Bulletin*, 112(1), S. 155-159
- Coleman, J. (1982): Systems of Trust. A rough theoretical framework. *Angewandte Sozialforschung*, 10, S. 277-299
- Coleman, J. (1990): Foundations of Social Theory. Cambridge, Mass.: Belknap Press of Harvard Univ. Press
- Coleman, J. (1991): Grundlagen der Sozialtheorie. Band 1: Handlungen und Handlungssysteme. München: Oldenbourg
- Coleman, J. (1994): Grundlagen der Sozialtheorie. Band 3: Die Mathematik der sozialen Handlung. München: Oldenbourg
- Comrey, A.L.; Lee, H.B. (1992): A first course in factor analysis. New York: Academic Press
- Cook, K.S.; Hardin, R.; Levi, M. (2005): Cooperation Without Trust? New York: Russell Sage Foundation

- Cooper, C.; Kline, P.; May, J. (1986): The Measurement of Authoritarianism, Psychoticism and other Traits by Objective Tests: A Cross-Validation. *Personality and Individual Differences*, 7(1), S. 15-21
- Cortina, J. M. (1993): What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78(1), S. 98-104
- Couch, L.; Adams, J.; Jones, W. (1996): The Assessment of Trust Orientation. *Journal of Personality Assessment*, 67(2), S. 305-323
- Couch, L.; Jones, W. (1997): Measuring Levels of Trust. *Journal of Research in Personality*. 31, S. 319-336
- Cronbach, L. J. (1951): Coefficient Alpha and the internal structure of tests. *Psychometrika*, 16(3), S. 297-334
- Cronbach, L.J. (1970): Essentials of psychological testing. New York: Harper & Row
- Dahl, G. (1971): Zur Berechnung des Schwierigkeitsindex bei quantitative abgestufter Aufgabenbewertung. *Diagnostica*, 17(1), S. 139-142
- Dasgupta, P. (1988): Trust as Commodity. In Gambetta, D. (Hrsg.): Trust. Making and Breaking Cooperative Relations. New York: Blackwell
- De Houwer, J. (2003): A structural analysis of indirect measures of attitudes. In: Musch, J.; Klauer, K.C. (Hrsg.): The Psychology of Evaluation. Affective Processes in Cognition and Emotion. New Jersey: Lawrence Erlbaum Associates, Inc.
- De Houwer, J. (2006): What are Implicit Measures and why are we using them? In: Wiers, R., W.; Stacy, A.W. (Hrsg.): Handbook of Implicit Cognition and Addiction. Thousand Oaks: Sage Publications
- De Houwer, J.; Moors, A. (2007): How to define and examine the implicitness of implicit measures. In: Wittenbrink, B.; Schwarz, N. (Hrsg.): Implicit Measures of Attitude. New York: Guilford Press
- Dederichs, A. (1997): Vertrauen als affektive Handlungsdimension: Ein emotionssoziologischer Bericht. In: Schweer, M. (Hrsg.): Vertrauen und soziales Handeln. Facetten eines alltäglichen Phänomens. Neuwied: Luchterhand
- Delhey, J.; Newton, K. (2002): Who trusts? The Origins of Social Trust in Seven Nations. Veröffentlichungen der Abteilung Sozialstruktur und Sozialberichterstattung des Forschungsschwerpunktes Sozialer Wandel, Institutionen und Vermittlungsprozesse, WZB, FS III 02-402
- Delhey, J.; Newton, K. (2004): Determinanten sozialen Vertrauens. Ein international vergleichender Theorietest. In: Klein, A.; Kern, K.; Geißel, B.; Berger, M. (Hrsg.): Zivilgesellschaft und Sozialkapital. Herausforderungen politischer und sozialer Integration. Wiesbaden: VS Verlag für Sozialwissenschaften
- Delhey, J.; Newton, K. (2005): Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism? *European Sociological Review*, 21(4), S. 311-327
- Delhey, J.; Newton, K.; Welzel, C. (2011): How General is Trust in "Most People"? Solving the Radius of Trust Problem. *American Sociological Review*, 76, S. 786-807
- Delhey, J.; Newton, K.; Welzel, C. (2014): The Radius of Trust Problem Remains Resolved. *American Sociological Review*, 79(6), S. 1260-1265

- Deutsch, M. (1960a): The Effect of Motivational Orientation upon Trust and Suspicion. *Human Relations*, 13, S. 123-139
- Deutsch, M. (1962): Cooperation and Trust: Some Theoretical Notes. In: Current Theory and Research in Motivation X: Nebraska Symposium on Motivation, 10, Lincoln/NE: University of Nebraska Press, S. 275-320
- Deutsch, M. (1976): Konfliktregelung. Konstruktive und destruktive Prozesse. München: Reinhardt
- Dieckmann, J. (2006): Schlüsselbegriffe der Systemtheorie. München: Fink
- Diekmann, A. (2011): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt
- Dislich, F.X.R.; Zinkernagel, A.; Ortner, T.M.; Schmitt, M. (2010): Convergence of Direct, Indirect, and Objective Risk-Taking Measures in Gambling. The Moderating Role of Impulsiveness and Self-Control. *Journal of Psychology*, 218(1), S. 20-27
- DIW Berlin/SOEP (Ed.) (2014): SOEP 2013 – Erhebungsinstrumente 2013 (Welle 30) des Sozio-oekonomischen Panels: Personenfragebogen, Altstichproben, SOEP Survey Papers, No. 180
- Döring, N.; Bortz, J. (2016): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Heidelberg: Springer
- Dürkheim, E. (1992): Über soziale Arbeitsteilung: Studien über die Organisation höherer Gesellschaften. Frankfurt am Main: Suhrkamp
- Dziuban, C. D.; Shirkey, E. C. (1974): When is a Correlation Matrix appropriate for Factor Analysis? Some Decision Rules. *Psychological Bulletin*, 81(6), S. 358-361
- Eid, M.; Gollwitzer, M.; Schmitt, M. (2010): Statistik und Forschungsmethoden. Weinheim: Beltz
- Eid, M.; Gollwitzer, M.; Schmitt, M. (2013): Statistik und Forschungsmethoden. Weinheim: Beltz
- Eid, M.; Gollwitzer, M.; Schmitt, M. (2015): Statistik und Forschungsmethoden. Weinheim: Beltz
- Elliot, S.; Lawty-Jones, M.; Jackson, C. (1996): Effect of Dissimulation on Self-Report and Objective Measures of Personality. *Personality and Individual Differences*, 21(3), S. 335-343
- Endress, M. (2002): Vertrauen. Bielefeld: transcript Verlag
- Endruweit, G.; Trommsdorff, G. (2002): Wörterbuch der Soziologie. Stuttgart: Lucius&Lucius
- Erikson, E. H. (1970): Jugend und Krise. Die Psychodynamik im sozialen Wandel. Stuttgart: Klett
- Erikson, E. H. (1974): Kindheit und Gesellschaft. Stuttgart: Klett
- Erlebacher, A. (1977): Design and Analysis of Experiments Contrasting the Within- and Between-Subjects Manipulation of the Independent Variable. *Psychological Bulletin*, 84(2), S. 212-219
- Esser, H. (1975a): Soziale Regelmäßigkeiten des Befragtenverhaltens. Meisenheim am Glan: Verlag Anton Hain
- Esser, H. (1975b): Zum Problem der Reaktivität bei Forschungskontakten. *KZfSS*, 27, S. 257-272

- Esser, H. (1986): Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *KZfSS*, 38, S. 314-336
- Esser, H. (1999): Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln. Frankfurt: Campus
- Eysenck, H.J. (1960a): Experiments in Personality. Volume II. Psychodiagnostics and Psychodynamics. London: Routledge and Kegan Paul
- Eysenck, H.J. (1960b): The Structure of Human Personality. London: Methuen
- Faas, T. (2009): Das Experiment – ein unbekanntes Wesen? In: Schnapp, K.-U.; Behnke, N.; Behnke, J. (Hrsg.): Datenwelten. Datenerhebung und Datenbestände in der Politikwissenschaft. Baden-Baden: Nomos
- Fahrenberg, J. (1964): Objektive Tests zur Messung der Persönlichkeit. In: Gottschaldt, K.; Lersch, P.; Heiß, R.; Bartenwerfer, H. (Hrsg.): Handbuch der Psychologie. Psychologische Diagnostik. Göttingen: Hogrefe
- Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A. (2007): G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), S. 175-191
- Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A. (2009): Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Research Methods*, 41(4), S. 1149-1160
- Fazio, R.H.; Olson, M.A. (2003): Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology*, 54, S. 297-327
- Fink, A. (1995): The survey kit. How to sample in surveys. Thousand Oaks: Sage
- Fischer, P.; Asal, K.; Krueger, J.I. (2014): Sozialpsychologie. Berlin: Springer
- Fisher, R.J. (1993): Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2), S. 303-315
- Fisher, R.J.; Tellis, G.J. (1998): Removing Social Desirability Bias with Indirect Questioning: Is the Cure Worse than the Disease? *Advances in Consumer Research*, 25(1), S. 563-567
- Fisseni, H. J. (1997): Lehrbuch der psychologischen Diagnostik. Göttingen: Hogrefe
- Fisseni, H. J. (2004): Lehrbuch der psychologischen Diagnostik. Göttingen: Hogrefe
- Forschungsdatenzentrum ALLBUS (2017): ALLBUS 2016: Fragebogendokumentation; ZA5250_fb.pdf
- Freitag, M.; Bühlmann, M. (2005): Politische Institutionen und die Entwicklung generalisierten Vertrauens. Ein internationaler Vergleich. *Politische Vierteljahresschrift*, 46(4), S. 575-601
- Freitag, M.; Bühlmann, M. (2009): Crafting Trust. The Role of Political Institutions in a Comparative Perspective. *Comparative Political Studies*, 42(12), S. 1537-1566
- Fritsche, I.; Linneweber, V. (2006): Nonreactive methods in psychological research. In: Eid, M.; Diener, E. (Hrsg.): Handbook of Multimethod Measurement in Psychology. Washington: American Psychological Association

- Fuhse, J.A.; Schaal, G.S. (2005): Politische Institutionen und die Generalisierung von Vertrauen. In: Geis, A.; Strecker, D. (Hrsg.): Blockaden staatlicher Politik. Sozialwissenschaftliche Analysen im Anschluss an Claus Offe. Frankfurt/Main: Campus Verlag
- Fukuyama, F. (2000): Social Capital and Civil Society. IMF Working Paper, IMF Institute, April 2000
- Fukuyama, F. (2001): Social capital, civil society and development. *Third World Quarterly*, 22(1), S. 7-20
- G*Power 3.1 manual (2007) http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf (Letzter Zugriff: 30.10.2017)
- Gabriel, O.W. (1999): Integration durch Institutionenvertrauen? Struktur und Entwicklung des Verhältnisses der Bevölkerung zum Parteienstaat und zum Rechtsstaat im vereinten Deutschland. In: Friedrichs, J.; Jagodzinski, W. (Hrsg.): Soziale Integration. *KZfSS*, Sonderheft 39, S. 199-235
- Gabriel, O.W. (2002): Politisches Vertrauen. In: Greiffenhagen, M.; Greiffenhagen, S. (Hrsg.): Handwörterbuch zur politischen Kultur in der Bundesrepublik Deutschland. Opladen: Westdeutscher Verlag, S. 494-499
- Gaines, B.J.; Kuklinski, J.H.; Quirk, P.J. (2007): The Logic of the Survey Experiment Reexamined. *Political Analysis*, 15(1), S. 1-20
- Gambetta, D. (2001): Können wir dem Vertrauen vertrauen? In: Hartmann, M.; Offe, C. (Hrsg.): Vertrauen. Die Grundlage des sozialen Zusammenhalts. Frankfurt: Campus-Verlag
- Garfinkel, H. (1963): A Conception of, and Experiments with, „Trust“ as a Condition of stable concerted Actions. In: Harvey, O. J. (Hrsg.): Motivation and Social Interaction. Cognitive Determinants. New York: The Ronald Press
- Garfinkel, H. (1967): Studies in Ethnomethodology. Englewood Cliffs, N.J.: Prentice-Hall
- Gawronski, B.; De Houwer, J. (2014): Implicit Measures in Social and Personality Psychology. In: Reis, H.T.; Judd, C.M. (Hrsg.): Handbook of Research Methods in Social and Personality Psychology. New York: Cambridge University Press
- Geis, F. (1978): Machiavellianism. In: London, H.; Exner, J. (Hrsg.): Dimensions of Personality. New York: John Wiley & Sons Inc.
- Giddens, A. (1992): The Consequences of Modernity. Stanford/Calif.: Stanford University Press
- Giddens, A. (1996a): Konsequenzen der Moderne. Frankfurt/Main: Suhrkamp Verlag
- Giddens, A. (1996b): Risiko, Vertrauen und Reflexivität. In: Beck, U.; Giddens, A.; Lash, S. (Hrsg.): Reflexive Modernisierung. Eine Kontroverse. Frankfurt/Main: Suhrkamp Verlag, S. 316-337
- Giddens, A. (1997a): Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung. Frankfurt/Main: Campus Verlag
- Giddens, A. (1997b): Jenseits von Links und Rechts. Die Zukunft radikaler Demokratie. Frankfurt/Main: Suhrkamp Verlag
- Glaeser, E; Laibson, D.; Scheinkman, J.; Soutter, C. (2000): Measuring Trust. *The Quarterly Journal of Economics*, 115, S. 811-846

- Göhler, G. (2002): Stufen des politischen Vertrauens. In: Schmalz-Bruns, R.; Zintl, R. (Hrsg.): *Politisches Vertrauen. Soziale Grundlagen reflexiver Kooperation*. Baden-Baden: Nomos
- Goodall, C. (2011): An overview of implicit measures of attitudes: Methods, mechanisms, strengths and limitations. *Communications Methods and Measures*, 5(3), S. 203-222
- Gordon, M. E.; Gross, R. H. (1978): A critique of methods for operationalizing the concept of fakeability. *Educational and Psychological Measurement*, 38(3), 771-782
- Gorsuch, R. L. (1983): *Factor Analysis*. Hillsdale: Erlbaum
- Govier, T. (1994): Is It a Jungle Out There? Trust, Distrust and the Construction of Social Reality. *Dialogue*, 33(2), S. 237-252
- Granovetter, M.S. (1973): The Strength of Weak Ties. *American Journal of Sociology*, 78(6), S. 1360-1380
- Greenwald, A. G. (1976): Within-Subjects Designs: To Use or Not To Use? *Psychological Bulletin*, 83(2), S. 314-320
- Greenwald, A.G.; Banaji, M.R. (1995): Implicit Cognition: Attitudes, Self-Esteem and Stereotypes. *Psychological Review*, 102(1), S. 4-27
- Guadagnoli, E.; Velicer, W. F. (1988): Relation of Sample Size to the Stability of Component Patterns. *Psychological Bulletin*, 103(2), S. 265-275
- Guilford, J. P. (1954): *Psychometric Methods*. New York: McGraw-Hill
- Gundelach, B. (2014a): *Ethnische Diversität und Soziales Vertrauen*. Baden-Baden: Nomos
- Gundelach, B. (2014b): In Diversity We Trust: The Positive Effect of Ethnic Diversity on Outgroup Trust. *Political Behavior*, 36, S. 125-142
- Gundelach, B.; Freitag, M. (2014): Neighbourhood Diversity and Social Trust: An Empirical Analysis of Interethnic Contact and Group-specific Effects. *Urban Studies*, 51(6), S. 1236-1256
- Gundelach, B.; Traunmüller, R. (2014): Beyond Generalised Trust: Norms of Reciprocity as an Alternative Form of Social Capital in an Assimilationist Integration Regime. *Political Studies*, 62, S. 596-617
- Häcker, H.; Schwenkmezger, P.; Utz, H. (1979): Über die Verfälschbarkeit von Persönlichkeitsfragebogen und Objektiven Persönlichkeitstests unter SD-Instruktion und in einer Auslesesituation. *Diagnostica*, 25(1), S. 7-23
- Häder, M. (2015): *Empirische Sozialforschung. Eine Einführung*. Wiesbaden: Springer
- Hardin, R. (1998): Trust in Government. In: Braithwaite, V; Levi, M. (Hrsg.): *Trust and Governance*. New York, NY: Russell Sage Foundation
- Hardin, R. (2001b): Die Alltagsepistemologie von Vertrauen. In: Hartmann, M.; Offe, C. (Hrsg.): *Vertrauen. Die Grundlage des sozialen Zusammenhalts*. Frankfurt: Campus-Verlag
- Hardin, R. (2002): *Trust and Trustworthiness*. New York: Russell Sage Foundation
- Hardin, R. (2004): *Distrust*. New York: Russell Sage Foundation
- Hardin, R. (2006): *Trust*. Cambridge: Polity Press

- Hartmann, P. (1991): Wunsch und Wirklichkeit. Theorie und Empirie sozialer Erwünschtheit. Wiesbaden: Deutscher Universitäts-Verlag
- Hefner, D. (2013): Implizite Methoden in der Kommunikationswissenschaft. In: Möhring, W.; Schlütz, D. (Hrsg.): Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft. Wiesbaden: Springer
- Heller, J. (2012): Experimentelle Psychologie. Eine Einführung. München: Oldenbourg Verlag
- Helliwell, J.F. (2003): How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling*, 20(2), S. 331-360
- Helliwell, J.F.; Putnam, R.D. (2004): The social context of well-being. *Philosophical Transactions of the Royal Society of London*, 359, S. 1435-1446
- Henrich, S. J.; Heine, S. J.; Norenzayan, A. (2010): The weirdest people in the world? *Behavioral and Brain Sciences*, 33, S. 61-135
- Herreros, F. (2004): The Problem of Forming Social Capital: Why Trust? New York: Palgrave Macmillan
- Herzberg, P.Y. (2011): Selbstdarstellung in Persönlichkeitsfragebögen: Das Phänomen der sozialen Erwünschtheit. In: Hornke, L.F.; Amelang, M.; Kersting, M. (Hrsg.): Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie II: Psychologische Diagnostik. Band 4: Persönlichkeitsdiagnostik. Göttingen: Hogrefe
- Herzberg, P.Y.; Roth, M. (2014): Persönlichkeitspsychologie. Wiesbaden: Springer VS
- Hirschle, J. (2015): Soziologische Methoden. Eine Einführung. Weinheim: Beltz Juventa
- Hofmann, K.; Kubinger, K.D. (2001): Herkömmliche Persönlichkeitsfragebogen und Objektive Persönlichkeitstests im „Wettstreit“ um (Un-)Verfälschbarkeit. *Report Psychologie*, 26, S. 298-304
- Huber, O. (2009): Das psychologische Experiment. Eine Einführung. Bern: Verlag Hans Huber
- Huck, S. W.; Sandler, H. M. (1973): A Note on the Solomon 4-Group Design: Appropriate Statistical Analyses. *The Journal of Experimental Education*, 42(2), S. 54-55
- Hundleby, J.D. (1973): The Measurement of Personality by Objective Tests. In: Kline, P. (Hrsg.): New Approaches in Psychological Measurement. London: Wiley
- Hussy, W.; Jain, A. (2002): Experimentelle Hypothesenprüfung in der Psychologie. Göttingen: Hogrefe
- Jasper, F.; Ortner, T. (2014): The Tendency to Fall for Distracting Information While Making Judgments. Development and Validation of the Objective Heuristic Thinking Test. *European Journal of Psychological Assessment*, 30(3), S. 193-207
- Jasper, F.; Ortner, T.M. (2014): The Tendency to Fall for Distracting Information while Making Judgements. Development and Validation of the Objective Heuristic Thinking Test. *European Journal of Psychological Assessment*, 30(3), S. 193-207
- Jonas, K.; Stroebe, W.; Hewstone, M. (2007): Sozialpsychologie. Eine Einführung. Heidelberg: Springer

- Jones, E.E.; Sigall, H. (1971): The Bogus Pipeline: A new Paradigm for Measuring Affect and Attitudes. *Psychological Bulletin*, 76(5), S. 349-364
- Jones, K. (1996): Trust as an Affective Attitude. *Ethics*, 107, S. 4-25
- Jones, K. (2013): Trust. In: LaFolette, H. (Hrsg.): The International Encyclopedia of Ethics. West Sussex: Blackwell Publishing
- Junge, K. (1998): Vertrauen und die Grundlage der Sozialtheorie. Ein Kommentar zu James S. Coleman. In: Müller H.-P.; Schmid, M. (Hrsg.): Norm, Herrschaft und Vertrauen. Beiträge zu James S. Colemans Grundlagen der Sozialtheorie. Opladen: Westdeutscher Verlag
- Kaiser, H. F.; Rice, J. (1974): Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34, S. 111-117
- Kaplan, R. (1973): Components of Trust: Note on Use of Rotter's Scale. *Psychological Reports*, 33, S. 13-14
- Kassebaum, U.B. (2004): Interpersonelles Vertrauen: Entwicklung eines Inventars zur Erfassung spezifischer Aspekte des Konstrukts. Hamburg: Dissertation
- Kirk, R. E. (2013): Experimental Design. Procedures for the Behavioral Sciences. Thousand Oaks: Sage Publications
- Kleining, G. (1986): Das qualitative Experiment. *KZfSS*, 38, S. 724-750
- Kline, P. (1979): Psychometrics and Psychology. London: Academic Press
- Kline, P.; Cooper, C. (1984): A Construct Validation of the Objective-Analytic Test Battery (OATB). *Personality and Individual Differences*, 5(2), S. 323-337
- Klopp, E. (2013): Explorative Faktorenanalyse. http://psydok.sulb.uni-saarland.de/volltexte/2013/4823/pdf/Explorative_Faktorenanalyse_final.pdf (Letzter Zugriff: 07.03.2016)
- Koch, T.; Ortner, T.M.; Eid, M.; Caspers, J.; Schmitt, M. (2014): Evaluating the Construct Validity of Objective Personality Tests Using a Multitrait-Multimethod-Multioccasion-(MTMM-MO)-Approach. *European Journal of Psychological Assessment*, 30(3), S. 208-230
- Korman, A. K. (1971). Industrial and organizational psychology. Englewood Cliffs: Prentice Hall.
- Kramer, R.M.; Brewer, M.B.; Hanna, B.A. (1996): Collective Trust and Collective Action. The Decision to Trust as a Social Decision. In: Kramer, R. M.; Tyler, T. R. (Hrsg.): Trust in Organizations. Frontiers of Theory and Research. Thousand Oaks: Sage
- Krampen, G.; Viebig, J.; Walter, W. (1982): Entwicklung einer Skala zur Erfassung dreier Aspekte von Sozialem Vertrauen. *Diagnostica*, 28(3), S. 242-247
- Kristof, W. (1983): Klassische Testtheorie und Testkonstruktion. In: Feger, H.; Bredenkamp, J. (Hrsg.): Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band 3: Messen und Testen. S. 544-603. Göttingen: Hogrefe
- Krohne, H.W.; Hock, M. (2015): Psychologische Diagnostik. Grundlagen und Anwendungsfelder. Stuttgart: Kohlhammer
- Krumpal, I. (2013): Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4), S. 2025-2047

- Kubinger, K.D. (1996): Objektive Diagnostik. In: Pawlik, K. (Hrsg.): Enzyklopädie der Psychologie. Themenbereich C: Theorie und Forschung. Serie VIII: Differentielle Psychologie und Persönlichkeitsforschung. Band 1: Grundlagen und Methoden der Differentiellen Psychologie. Göttingen: Hogrefe
- Kubinger, K.D. (1997): Zur Renaissance der objektiven Persönlichkeitstests sensu R.B. Cattell. In: Mandl, H.; Bruckmoser, S.; Gruber, H. (Hrsg.): Schwerpunktthema Wissen und Handeln. Bericht über den 40. Kongress der deutschen Gesellschaft für Psychologie. Göttingen: Hogrefe
- Kubinger, K.D. (2005): Objektive psychologisch-diagnostische Verfahren. In: Weber, H.; Ramm-sayer, T. (Hrsg.): Handbuch der Persönlichkeitspsychologie und differentiellen Psychologie. Göttingen: Hogrefe
- Kubinger, K.D. (2006): Ein Update der Definition von Objektiven Persönlichkeitstests: Experimentalpsychologische Verhaltensdiagnostik. In: Ortner, T.M.; Proyer, R.T.; Kubinger, K.D. (Hrsg.): Theorie und Praxis objektiver Persönlichkeitstests. Bern: Huber
- Kubinger, K.D. (2009): The technique of objective personality-tests sensu R.B. Cattell nowadays: The Viennese pool of computerized tests aimed at experiment-based assessment of behavior. *Acta Psychologica Sinica*, 41, S. 1024-1036
- Kubinger, K.D.; Rasch, D.; Moder, K. (2009): Zur Legende der Voraussetzungen des t-Tests für unabhängige Stichproben. *Psychologische Rundschau*, 60, S. 26-27
- Kuckartz, U.; Rädiker, S.; Ebert, T.; Schehl, J. (2013): Statistik. Eine verständliche Einführung. Wiesbaden: Springer VS
- Kunz, V. (2004): Soziales Vertrauen. In: van Deth, J.W. (Hrsg.): Deutschland in Europa. Ergebnisse des European Social Survey 2002-2003. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lahno, B. (1999): Ist Vertrauen eine rationale Erwartung? In: Nida-Rümelin, J. (Hrsg.): Rationalität, Realismus, Revision. Vorträge des 3. Internationalen Kongresses der Gesellschaft für Analytische Philosophie vom 15. Bis zum 18. September 1997 in München. Berlin: de Gruyter
- Lahno, B. (2000): Verbundenheit als konstitutives Element des Vertrauens. In: Beckmann, K. (Hrsg.): Individuum versus Kollektiv: der Kommunitarismus als "Zauberformel"? Frankfurt am Main: Lang
- Lahno, B. (2001): On the Emotional Character of Trust. *Ethical Theory and Moral Practice*, 4, S. 171-189
- Lahno, B. (2002): Der Begriff des Vertrauens. Paderborn: Mentis
- Lahno, B. (2004): Three Aspects of Interpersonal Trust. *Analyse&Kritik*, 26, S. 30-47
- Lahno, B. (2005): Vertrauen. In: Mummert, U.; Sell, F. (Hrsg.): Emotionen, Markt und Moral. Münster: LIT Verlag
- Lahno, B. (2011): Trust. In: Dowding, K. (Hrsg.): Encyclopedia of Power. California: Sage Publications
- Lamla, J. (2003): Anthony Giddens. Frankfurt/Main: Campus Verlag
- Lamnek, S. (2010): Qualitative Sozialforschung. Weinheim: Beltz

- Lamp, E. (2001): Ist einer von drei gleich jedem Dritten? Der Einfluss numerischer Äquivalente auf die Wahrnehmung und Bewertung identischer Sachverhalte. *ZA-Information/Zentralarchiv für Empirische Sozialforschung*, 49, S. 49-68
- Lee, R.M. (1993): *Doing Research on Sensitive Topics*. London: SAGE Publications
- Lee, R.M. (2000): *Unobtrusive methods in social research*. Philadelphia: Open University Press
- Leichsenring, F.; Hiller, W. (1994): Projektive Verfahren. In: Stieglitz, R.-D.; Baumann, U. (Hrsg.): *Psychodiagnostik psychischer Störungen*. Stuttgart: Enke
- Leiner, D. J. (2014): Convenience Samples from Online Respondent Pools: A case study of the SoSci Panel. Working Paper. https://www.researchgate.net/publication/259669050_Convenience_Samples_from_Online_Respondent_Pools_A_case_study_of_the_SoSci_Panel (Letzter Zugriff: 07.04.2016)
- Lejeuz, C.W.; Read, J.P.; Kahler, C.W.; Richards, J.B.; Ramsey, J.B.; Stuart, G.L.; Strong, D.R.; Brown, R.A. (2002): Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology-Applied*, 8, S. 75-84
- Lejeuz, C.W.; Read, J.P.; Kahler, C.W.; Richards, J.B.; Ramsey, S.E.; Stuart, G.L.; Strong, D.R.; Brown, R.A. (2002): Evaluation of a Behavioral Measure of Risk Taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), S. 75-84
- Lemon, N. (1973): *Attitudes and Their Measurement*. London: B.T. Batsford Ltd.
- Lenzner, T.; Neuert, C.; Otto, W. (2014): *Kognitive Pretesting. SDM Survey Guidelines*, Mannheim: GESIS-Leibniz Institut für Sozialforschung
- Levi, M. (1996): Social and Unsocial Capital: A Review Essay of Robert Putnam's *Making Democracy Work*. *Politics & Society*, 24(1), S. 45-55
- Levi, M. (1998): A State of Trust. In: Braithwaite, V.; Levi, M. (Hrsg.): *Trust and Governance*. New York: Russell Sage Foundation
- Levi, M. (2000): When good defenses make good neighbors: a transaction cost approach to trust, the absence of trust and distrust. In: Ménard, C. (Hrsg.): *Institutions, Contracts and Organizations. Perspectives from New Institutional Economics*. Cheltenham: Edward Elgar
- Levi, M. (2001): Sociology of Trust. In: Smelser, N.; Baltes, P. (Hrsg.): *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier
- Levi, M.; Stoker, M. (2000): Political Trust and Trustworthiness. *Annual Review of Political Science*, 3, S. 475-507
- Lewicki, R.; Bunker, B. (1995): Trust in Relationships. A Model of Development and Decline. In: Bunker, B.; Rubin, J. (Hrsg.): *Conflict, Cooperation, and Justice. Essays Inspired by the Work of Morton Deutsch*. San Francisco: Jossey-Bass Publishers
- Lewicki, R.; Bunker, B. (1996): Developing and Maintaining Trust in Work Relationships. In: Kramer, R. M.; Tyler, T.R. (Hrsg.): *Trust in Organizations. Frontiers of Theory and Research*. Thousand Oaks: Sage
- Lewis, J.D.; Weigert, A.J. (1985a): Trust as a Social Reality. *Social Forces*, 63(4), S. 967-985

- Lewis, J.D.; Weigert, A.J. (1985b): Social Atomism, Holism, and Trust. *Sociological Quarterly*, 26(4), S. 455-471
- Lienert, G. A.; Raatz, U. (1998): Testaufbau und Testanalyse. Weinheim: Beltz
- Loevinger, J. (1957): Objective Tests as Instruments of Psychological Theory. *Psychological Reports*, 3, S. 635-694
- Lucas, R.E.; Baird, B.M. (2006): Global Self-Assessment. In: Eid, M.; Diener, E. (Hrsg.): Handbook of Multimethod Measurement in Psychology. Washington: American Psychological Association
- Lück, H.; Timaeus, E. (1969): Skalen zur Messung manifester Angst (MAS) und sozialer Wünschbarkeit (SDS-E und SDS-CM). *Diagnostica*, 15, S. 134-141
- Lück, H.; Timaeus, E. (2014): Soziale Erwünschtheit (SDS-CM). In: Danner, D.; Glöckner-Rist, A. (Hrsg.): Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 15.00
- Luhmann, N. (1979): Trust and Power. Chichester: Wiley & Sons Ltd.
- Luhmann, N. (1987): Soziale Systeme. Grundriß einer allgemeinen Theorie. Frankfurt am Main: Suhrkamp
- Luhmann, N. (1988): Familiarity, Confidence, Trust: Problems and Alternatives. In: Gambetta, D. (Hrsg.): Trust. Making and Breaking Cooperative Relations. New York: Blackwell, S. 94-107
- Luhmann, N. (2001): Vertrautheit, Zuversicht, Vertrauen: Probleme und Alternativen. In: Hartmann, M.; Offe, C. (Hrsg.): Vertrauen. Die Grundlage des sozialen Zusammenhalts. Frankfurt am Main: Campus-Verlag, S. 143-160
- Luhmann, N. (2009): Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität. Stuttgart: Lucius & Lucius
- MacCallum, R. C.; Widaman, K. F.; Zhang, S.; Hong, S. (1999): Sample Size in Factor Analysis. *Psychological Methods* 4(1), S. 84-99
- Maddux, W. W.; Brewer, M. B. (2005): Gender Differences in the Relational and Collective Bases for Trust. *Group Processes & Intergroup Relations*, 8(2), S. 159-171
- Mayer, R.; Davis, J.; Schoorman, F. (1995): An integrative Model of Organizational Trust. *Academy of Management Review*, 20(3), S. 709-734
- Mayrhofer, W. (1993): Nonreaktive Methoden. In: Becker, F.G.; Martin, A. (Hrsg.): Empirische Personalforschung. Methoden und Beispiele. München: Rainer Hampp Verlag
- Mendoza, J. L.; Stafford, K. L.; Stauffer, J. M. (2000): Large-Sample Confidence Intervals for Validity and Reliability Coefficients. *Psychological Methods*, 5(3), S. 356-369
- Merten, K. (2005): Reaktivität und Reflexivität: Sozialwissenschaftliche Datenerhebung als interferierende Kommunikationsprozesse. In: Wienand, E.; Westerbarkey, J.; Scholl, A. (Hrsg.): Kommunikation über Kommunikation. Theorien, Methoden und Praxis. Festschrift für Klaus Merten. Wiesbaden: VS-Verlag
- Meschede, D. (2002): Gerthsen Physik. Berlin: Springer-Verlag
- Messick, D. M.; Kramer, R. M. (2001): Trust as a Form of Shallow Morality. In: Cook, K. (Hrsg.): Trust in Society. New York: Russell Sage Foundation

- Meyer, G.J.; Kurtz, J.E. (2006): Advancing Personality Assessment Terminology: Time to Retire "Objective" and "Projective" As Personality Test Descriptors. *Journal of Personality Assessment*, 87(3), S. 223-225
- Meyer, H. (2004): Theorie und Qualitätsbeurteilung psychometrischer Tests. Stuttgart: Kohlhammer
- Meyerson, D.; Weick, K. E.; Kramer, R. M. (1996): Swift Trust and Temporary Groups. In: Kramer, R. M.; Tyler, T. R. (Hrsg.): Trust in Organizations. Frontiers of Theory and Research. Thousand Oaks: Sage
- Miller, A.; Mitamura, T. (2003): Are Surveys on Trust Trustworthy. *Social Psychology Quarterly*, 66(1), S. 62-70
- Miller, G. (1956): The magic number 7, plus or minus two: some limits on our capacity for processing information. *Psychological Reports*, 63, S. 81-97
- Miller, J. (2000): Trust: The Moral Importance of an Emotional Attitude. *Practical Philosophy*, 3(3), S. 45-54
- Miller, J. (2002): A Critical Moral Ethnography of Social Distrust. *Social Philosophy Today*, 16, S. 141-158
- Miller, J. (2003): Trust in Strangers, Trust in Friends. *Philosophy in the Contemporary World*, 10(1), S. 17-22
- Miller, J.D. (1985): The Nominative Technique: A new Method of Estimating Heroin Prevalence. In: Rouse, B.A.; Kozel, N.J.; Richards, L.G. (Hrsg.): Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity. NIDA Research Monograph, 57, S. 104-124
- Misztal, B. (1996): Trust in Modern Societies. The Search for the Bases of Social Order. Cambridge: Polity Press
- Möhring, W.; Schlütz, D. (2010): Die Befragung in der Medien- und Kommunikationswissenschaft. Eine praxisorientierte Einführung. Wiesbaden: VS-Verlag
- Moosbrugger, H.; Kelava, A. (2012): Testtheorie und Fragebogenkonstruktion. Berlin: Springer
- Mortimer, C.E.; Müller, U. (2001): Chemie. Das Basiswissen der Chemie. Stuttgart: Georg Thieme Verlag
- Mühlenfeld, H.-S. (2004): Der Mensch in der Online-Kommunikation. Zum Einfluss webbasierter, audiovisueller Fernkommunikation auf das Verhalten von Befragten. Wiesbaden: Deutscher Universitäts Verlag
- Mummendey (1981): Methoden und Probleme der Kontrolle sozialer Erwünschtheit (Social Desirability). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2(3), S. 199-218
- Mummendey, H. D.; Grau, I. (2014): Die Fragebogen-Methode. Grundlagen und Anwendungen in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung. Göttingen: Hogrefe
- Mummendey, H.D.; Eifler, S. (1993): Eine neue Skala zur Messung sozialer Erwünschtheit. Bielefelder Arbeiten zur Sozialpsychologie Nr. 167.
- Mundfrom, D. J.; Shaw, D. G.; Ke T. L. (2005): Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing*, 5(2), S. 159-168

- Musch, J.; Brockhaus, R.; Bröder, A. (2002): Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica*, 48(3), S. 121-129
- Naef, M.; Schupp, J. (2009): Measuring Trust: Experiments and Surveys in Contrast and Combination. IZA Discussion Paper No. 4087, March 2009
- Nannestad, P. (2008): What have we learned about generalized trust, if anything? *Annual Review of Political Science*, 11, S. 413-436
- Newton, K. (1999a): Social and Political Trust in established Democracies. In: Norris, P. (Hrsg.): *Critical Citizens. Global Support for democratic Government*. Oxford, Oxford University Press
- Newton, K. (1999b): Social capital and democracy in modern Europe. In: van Deth, J.; Maraffi, M.; Whiteley, P. (Hrsg.): *Social Capital and European Democracy*. London, Routledge
- Newton, K. (2001): Trust, Social Capital, Civil Society, and Democracy. *International Political Science Review*, 22(2), S. 201-214
- Newton, K. (2008): Trust and Politics. In: Castiglione, D.; van Deth, J.; Wolleb, G. (Hrsg.): *The Handbook of Social Capital*. Oxford, Oxford University Press
- Newton, K.; Norris, P. (2000): Confidence in Public Institutions: Faith, Culture, or Performance? In: Pharr, S.; Putnam, R. (Hrsg.): *Disaffected Democracies. What's troubling the trilateral Countries?* Princeton, NJ: Princeton University Press
- Newton, K.; Zmerli, S. (2011): Three forms of trust and their association. *European Political Science Review*, 3(2), S. 169-200
- Nock, S.L.; Guterbock, T.M. (2010): Survey Experiments. In: Marsden, P.V.; Wright, J.D. (Hrsg.): *Handbook for Survey Research*. Bingley: Emerald
- Noll, H.-H. (2000): Subjektive Indikatoren. Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Zentrum für Umfragen, Methoden und Analysen (ZUMA)
- Nosek, B.A. (2007): Implicit-Explicit Relations. *Current Directions in Psychological Science*, 16(2), S. 65-69
- Nosek, B.A.; Greewald, A.G.; Banaji, M.R. (2007): The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In: Bargh, J.A. (Hrsg.): *Automatic processes in social thinking and behavior*. Psychology Press
- O'Connor, B. P. (2000): SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers*, 32(3), S. 396-402
- Öberg, P.; Oskarsson, S.; Svensson, T. (2011): Similarity vs. homogeneity: contextual effects in explaining trust. *European Political Science Review*, 3(3), S. 345-369
- Offe, C. (2000): Demokratie und Vertrauen. In: *Transit*, 18 (Winter 1999/2000): 1989 und die Folgen, S. 118-131
- Offe, C. (2001a): Wie können wir unseren Mitbürgern vertrauen? In: Hartmann, M.; Offe, C. (Hrsg.): *Vertrauen. Die Grundlage des sozialen Zusammenhalts*. Frankfurt: Campus-Verlag

- Offe, C. (2001b): Political Liberalism, Group Rights, and the Politics of Fear and Trust. *Studies in East European Thought*, 53, S. 167-182
- Orne, M. (1962): On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), S. 776-783
- Orne, M.T. (1969): Demand Characteristics and the Concept of Quasi-controls. In: Rosenthal, R.; Rosnow, R. L. (Hrsg.): *Artifact in Behavioral Research*. New York: Academic Press
- Ortner, T.M.; Horn, R.; Kersting, M.; Krumm, S. Kubinger, K.D. Proyer, R.T. Schmidt-Atzert, L. Schuhfried, G.; Schütz, A.; Wagner-Menghin, M. M.; Westhoff, K. (2007): Standortbestimmung und Zukunft Objektiver Persönlichkeitstests. *Report Psychologie*, 32, S. 60-69
- Ortner, T.M.; Proyer, R.T. (2015): Objective Personality Tests. In: Ortner, T.M.; van de Vijver, F.J.R. (Hrsg.): *Behavior-Based Assessment in Psychology. Going Beyond Self-Report in the Personality, Affective, Motivation and Social Domains*. Göttingen: Hogrefe
- Ortner, T.M.; Schmitt, M. (2014): Advances and Continuing Challenges in Objective Personality Testing. *European Journal of Psychological Assessment*, 30(3), S. 163-168
- Oskarsson, S.; Öberg, P.; Svensson, T. (2009): Making Capitalism Work: Fair Institutions and Trust. *Economic and industrial Democracy*, 30(2), S. 294-320
- Oskarsson, S.; Svensson, T.; Öberg, P.; (2007): Creating trust. The role of power asymmetries and institutional constraints. The Department of Government Studies in Political Economy and Welfare. Working Paper Series. Number 5/2007
- Oskarsson, S.; Svensson, T.; Öberg, P.; (2009): Power, Trust and Institutional Constraints. Individual Level Evidence. *Rationality and Society*, 21(2), S. 171-195
- Oswald, M. (2006): Vertrauen in Personen und Organisationen. In: Bierhoff, H.-W.; Frey, D. (Hrsg.): *Handbuch der Sozialpsychologie und Kommunikationspsychologie*. Band 3. Göttingen: Hogrefe
- Park, C.-M. (2012): Assessing Three Theories of Generalized Trust: Evidence from East Asia. Prepared for delivery at the 2012 Annual Meeting of the American Political Science Association, August 30-September 2, 2012
- Patzelt, W. J. (1987): Grundlagen der Ethnomethodologie. Theorie, Empirie und politikwissenschaftlicher Nutzen einer Soziologie des Alltags. München: Fink
- Paulhus, D. (1984): Two-Component Models of Socially Desirable Responding. *Journal of Personality and Social Psychology*, 46(3), S. 598-609
- Paulhus, D. (1986): Self-Deception and Impression Management in Test Responses. In: Angleitner, A.; Wiggins, J.S. (Hrsg.): *Personality Assessment via Questionnaires. Current Issues in Theory and Measurement*. Heidelberg: Springer-Verlag
- Pauls, C.A.; Stemmler, G. (2003): Substance and bias in social desirability responding. *Personality and Individual Differences*, 35, S. 263-275
- Pawlik, K. (2006): Objektive Tests in der Persönlichkeitsforschung. In: Ortner, T.M.; Proyer, R.T.; Kubinger, K.D. (Hrsg.): *Theorie und Praxis objektiver Persönlichkeitstests*. Bern: Huber
- Petermann, F. (2013): *Psychologie des Vertrauens*. Göttingen: Hogrefe Verlag

- Petermann, F.; Noack, H. (1999): Nicht-reaktive Meßverfahren. In: Roth, E.; Heidenreich, K.; Hölting, H. (Hrsg.): Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis. München: Oldenbourg
- Petersen, T. (2002): Das Feldexperiment in der Umfrageforschung. Frankfurt: Campus
- Petty, R.E.; Cacioppo, J.T. (1996): Attitudes and Persuasion: Classic and Contemporary Approaches. Boulder: Westview Press
- Petty, R.E.; Fazio, R.H.; Briñol, P. (2009): The new implicit measures. An overview. In: Petty, R.E.; Fazio, R.H.; Briñol, P. (Hrsg.): Attitudes. Insights from the New Implicit Measures. New York: Psychology Press
- Pospeschill, M. (2006): Statistische Methoden. Strukturen, Grundlagen, Anwendungen in Psychologie und Sozialwissenschaften. München: Elsevier
- Preisendörfer, P. (1995): Vertrauen als soziologische Kategorie. Möglichkeiten und Grenzen einer entscheidungstheoretischen Fundierung des Vertrauenskonzepts. *Zeitschrift für Soziologie*, 24(4), S. 263-272
- Proyer, R. (2007): Convergence of conventional and behavior-based measures: Towards a multimethod approach in the assessment of vocational interests. *Psychology Science*, 49(2), S. 168-183
- Proyer, R.T.; Häusler, J. (2007): Assessing behavior in standardized settings: The role of objective personality tests. *International Journal of Clinical and Health Psychology*, 7(2), S. 537-546
- Prüfer, P.; Rexroth, M. (2000): Zwei-Phasen-Pretesting. *ZUMA-Arbeitsbericht* 2000/08
- Putnam, R.D. (1995a): Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America. *Political Science and Politics*, 28(4), S. 664-683
- Putnam, R.D. (1995b): Bowling Alone, Revisited. *The Responsive Community*, (Spring), S. 18-33
- Putnam, R.D. (2000): Bowling Alone. The collapse and revival of American Community. New York: Simon & Schuster
- Putnam, R.D. (2007): E Pluribus Unum: Diversity and Community in the Twenty-first Century. The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies*, 30(2), S. 137-174
- Putnam, R.D.; Leonardi, R.; Nanetti, R. (1993): Making Democracy work. Civic traditions in modern Italy. Princeton, NJ: Princeton University Press
- Rahn, W.M.; Trensue, J.E. (1998): Social Trust and and Value Change: The Decline of Social Capital in American Youth, 1976-1995. *Political Psychology*, 19, S. 545-565
- Ramsey, P.H. (1980): Exact type I error rates for robustness of Student's t-test with unequal variances. *Journal of Educational Statistics*, 5(4), S. 337-349
- Rasch, D.; Guiard, V. (2004): The robustness of parametric statistical methods. *Psychology Science*, 46(2), S. 175-208
- Rasch, D.; Kubinger, K.D. (2006): Statistik. Für das Psychologiestudium. München: Elsevier
- Rauthmann, J.F. (2017): Persönlichkeitspsychologie. Paradigmen, Strömungen, Theorien. Berlin: Springer

- Reeskens, T.; Hooghe, M. (2008): Cross-cultural measurement equivalence of generalized trust. Evidence from the European Social Survey (2002 and 2004). *Social Indicators Research*, 85, S. 515-532
- Reinecke, J. (1991): Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte. Opladen: Westdeutscher Verlag
- Reiß, S.; Sarris, V. (2012): Experimentelle Psychologie. Von der Theorie zur Praxis. München: Pearson
- Reuband, K.-H. (1990): Meinungslosigkeit im Interview. Erscheinungsformen und Folgen unterschiedlicher Befragungsstrategien. *KZfSS*, 19(6), S. 428-443
- Revenstorf, D. (1976): Lehrbuch der Faktorenanalyse. Stuttgart: Kohlhammer
- Robinson, J.P.; Shaver, P.R., Wrightsman, L.S. (1991): Measures of Personality and Social Psychological Attitudes. San Diego: Academic Press
- Robinson, M.D.; Neighbors, C. (2006): Catching the mind in action: Implicit methods in personality research and assessment. In: Eid, M.; Diener, E. (Hrsg.): Handbook of Multimethod Measurement in Psychology. Washington: American Psychological Association
- Rose-Ackerman, S. (2001): Trust, honesty and corruption: Reflection on the state-building process. *European Journal of Sociology*, 42(3), S. 526-570
- Rosenberg, M. (1956): Misanthropy and Political Ideology. *American Sociological Review*, 21, S. 690-695
- Rosenberg, M. (1957a): Occupations and Values. Glencoe: The Free Press
- Rosenberg, M. (1957b): Misanthropy and attitudes toward international affairs. *Journal of Conflict Resolution*, 1, S. 340-345
- Rosenberg, M. (1965): Society and the Adolescent Self-Image. Princeton: Princeton University Press
- Rosenthal, R.; Rosnow, R. L. (2008): Essentials of Behavioral Research. Methods and Data Analysis. New York: McGraw-Hill
- Rost, J. (2004): Testtheorie – Testkonstruktion. Bern: Huber
- Rothstein, B.; Stolle, D (2001): Social Capital and Street-Level Bureaucracy: An Institutional Theory of Generalized Trust. Prepared for the "Trust in Government Conference" at the Center for the Study of Democratic Politics, Princeton University, November 30, 2001
- Rothstein, B.; Stolle, D (2002): How Political Institutions create and destroy Social Capital: An Institutional Theory of Generalized Trust. Paper to be presented at Collegium Budapest, Project on Honesty and Trust: Theory and Experience in the Light of Post-Socialist Experience Workshop 2: Formal and Informal Cooperation, November 22-23, 2002
- Rothstein, B.; Stolle, D (2003): Social Capital, Impartiality and the Welfare State: An Institutional Approach. In: Hooghe, M.; Stolle, D. (Hrsg.): Generating Social Capital. Civil Society and Institutions in Comparative Perspective. New York: Palgrave Macmillan

- Rothstein, B.; Stolle, D (2007): The Quality of Government and Social Capital: A theory of Political Institutions and Generalized Trust. The Quality of Government Institute: QoG Working Paper Series 2007: 2
- Rothstein, B.; Stolle, D (2008a): Political Institutions and Generalized Trust. In: v. Deth, J.; Castiglione, D.; Wolleb, G. (Hrsg.): The Handbook of Social Capital. Oxford: Oxford University Press
- Rothstein, B.; Stolle, D (2008b): The State and Social Capital. An Institutional Theory of Generalized Trust. *Comparative Politics*, 40, S. 441-459
- Rothstein, B.; Uslaner, E. M. (2005): All For All. Equality, Corruption, and Social Trust. *World Politics*, 58, S. 41-72
- Rotter, J. (1967): A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, S. 651-665
- Rotter, J. (1971): Generalized Expectancies for interpersonal Trust. *American Psychologist*, 26, S. 443-452
- Rotter, J. (1980): Interpersonal Trust, Trustworthiness, and Gullibility. *American Psychologist*, 35(1), S. 1-7
- Rotter, J. (1981): Vertrauen. *Psychologie heute*, 8(3), S. 23-29
- Rubio, V.J.; Hernández, J.M.; Zaldívar, F.; Márquez, O.; Santacreu, J. (2010): Can We Predict Risk-Taking Behavior? Two Behavioral Tests for Predicting Guessing Tendencies in a Multiple-Choice Test. *European Journal of Psychological Assessment*, 26(2), S. 87-94
- Santacreu, J.; Rubio, V.J.; Hernández, J.M. (2006): The objective assessment of personality: Cattell's T-Data revisited and more. *Psychology Science*, 48(1), S. 53-68
- Saris, W. E.; Satorra, A.; Coenders, G. (2004): A new approach to evaluating the quality of measurement instruments: The split-ballot MTMM Design. *Sociological Methodology*, 34(1), S. 311-347
- Schandry, R. (1996): Psychophysiologische Erhebungsmethoden. In: Pawlik, K. (Hrsg.): Enzyklopädie der Psychologie. Themenbereich C: Theorie und Forschung. Serie VIII: Differentielle Psychologie und Persönlichkeitsforschung. Band 1: Grundlagen und Methoden der Differentiellen Psychologie. Göttingen: Hogrefe
- Schmidt, L. (1975): Objektive Persönlichkeitsmessung in diagnostischer und klinischer Psychologie. Weinheim: Beltz
- Schmidt, L.R. (2006): Objektive Persönlichkeitstests in der Tradition Cattells: Forschungslinien und Relativierungen. In: Ortner, T.M.; Proyer, R.T.; Kubinger, K.D. (Hrsg.): Theorie und Praxis objektiver Persönlichkeitstests. Bern: Huber
- Schmidt, L.R.; Schwenkmezger, P. (1994): Objektive Persönlichkeitstests: Perspektiven für die Diagnostik. In: Bartussek, D.; Amelang, M. (Hrsg.): Fortschritte der Differentiellen Psychologie und Psychologischen Diagnostik. Göttingen: Hogrefe
- Schmitt, N. (1996): Uses and Abuses of Coefficient Alpha. *Psychological Assessment*, 8(4), S. 350-353

- Schmukle, S.C.; Egloff, B. (2011): Indirekte Verfahren zur Erfassung von Persönlichkeit ("Objektive Persönlichkeitstests"). In: Hornke, L.F.; Amelang, M.; Kersting, M. (Hrsg.): Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie II: Psychologische Diagnostik. Band 4: Persönlichkeitsdiagnostik. Göttingen: Hogrefe
- Schnell, R.; Hill, P.B.; Esser, E. (2008): Methoden der empirischen Sozialforschung. München: Oldenbourg
- Schnell, R.; Hill, P.B.; Esser, E. (2011): Methoden der empirischen Sozialforschung. München: Oldenbourg
- Scholl, A. (1993): Die Befragung als Kommunikationssituation. Zur Reaktivität im Forschungsinterview. Opladen: Westdeutscher Verlag
- Scholl, A. (2013): Reaktivität im Forschungsprozess. In: Möhring, W.; Schlütz, D. (Hrsg.): Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft. Wiesbaden: Springer VS
- Scholl, A. (2015): Die Befragung. Konstanz: UVK Verlagsgesellschaft
- Schumann, H.; Presser, S. (1981): Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording and Content. New York: Academic Press
- Schwartz, B. (2003): The Paradox of Choice: Why More is Less. New York: Ecco
- Schwarz, N. (1999): Self-Reports. How the questions shape the answers. *American Psychologist*, 54(2), S. 93-105
- Schweer, K.W. (2000): Politisches Vertrauen: Theoretische Ansätze und empirische Befunde. In: Schweer, K.W. (Hrsg.): Politische Vertrauenskrise in Deutschland? Eine Bestandsaufnahme. Münster: Waxmann
- Schweer, M. (1997): Eine differentielle Theorie interpersonalen Vertrauens. Überlegungen zur Vertrauensbeziehung zwischen Lehrenden und Lernenden. *Psychologie in Erziehung und Unterricht*, 44, S. 2-12
- Sechrest, L.; Phillips, M. (1979): Unobtrusive Measures: An Overview. In: Sechrest, L. (Hrsg.): Unobtrusive Measurement Today. San Francisco: Jossey-Bass Inc.
- Sedlmeier, P.; Renkewitz, F. (2008): Forschungsmethoden und Statistik in der Psychologie. München: Pearson
- Sedlmeier, P.; Renkewitz, F. (2013): Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler. München: Pearson
- Seligman, A. (1997): The Problem of Trust. Princeton, N.J: Princeton University Press
- Seligman, A. (1998): Trust and Sociability: On the Limits of Confidence and Role Expectations. *American Journal of Economics and Sociology*, 57(4), S. 391-404
- Seligman, A. (2000): Trust, Confidence and the Problem of Civility. In: Rouner, L. (Hrsg.): Civility. Notre Dame: University of Notre Dame Press
- Selltiz, C.; Jahoda, M.; Deutsch, M.; Cook, S.W. (1972): Untersuchungsmethoden der Sozialforschung Teil II. Darmstadt: Luchterhand

- Shadish, W. R.; Cook, T. D.; Campbell, D. T. (2002): *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin
- Shapiro, D.; Sheppard, B.; Cheraskin, L. (1992): Business on a Handshake. *Negotiation Journal*, 8(4), S. 365-377
- Siegel, S.; Castellan, N. J. (1988): *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill
- Simmel, G. (1989): Philosophie des Geldes. In: Frisby, D.P.; Köhnke, K.D. (Hrsg.): *Gesamtausgabe*, Band 6, Frankfurt am Main: Suhrkamp
- Simmel, G. (1992): Soziologie. Untersuchungen über die Formen der Vergesellschaftung. In: Rammstedt, O. (Hrsg.): *Gesamtausgabe*, Band 11, Frankfurt am Main: Suhrkamp
- Simon, R.J.; Simon, J.L. (1974): Money Incentives and Family Size: A Hypothetical-Question Study. *Public Opinion Quarterly*, 38(4), S. 585-595
- Sirken, M. (1970): Household Surveys with Multiplicity. *Journal of American Statistical Association*, 65(329), S. 257-266
- Smith, T.W. (1997): Factors Relating to Misanthropy in Contemporary American Society. *Social Science Research*, 26.S. 170-196
- Sniderman, P.M.; Grob, D.B. (1996): Innovations in Experimental Design in Attitude Surveys. *Annual Review of Sociology*, 22, S. 377-399
- Spector, P. E. (1992): *Summated Rating Scale Construction. An Introduction*. Newbury Park: Sage
- St. John, F.A.V.; Edwards-Jones, G.; Gibbons, J.M.; Jones, J.P.G. (2010): Testing novel methods for assessing rule breaking in conservation. *Biological Conservation*, 143, S. 1025-1030
- Stolle, D. (1998): Bowling Together, Bowling Alone: The Development of Generalized Trust in Voluntary Associations. *Political Psychology*, 19(3), S. 497-525
- Stolle, D. (2001a): Clubs and Congregations: The Benefits of Joining an Association. In: Cook, K. (Hrsg.): *Trust in Society*. New York: Russell Sage Foundation
- Stolle, D. (2001b): Getting to trust. An analysis of the importance of institutions, families, personal experiences and group membership. In: Dekker, P.; Uslaner, E. (Hrsg.): *Social Capital and Participation in Everyday Life*. London: Routledge
- Stolle, D. (2002): Trusting Strangers. The Concept of Generalized Trust in Perspective. *ÖZP*, 31 (4), S. 397-412
- Stolle, D. (2003): The Sources of Social Capital. In: Hooghe, M.; Stolle, D. (Hrsg.): *Generating Social Capital. Civil Society and Institutions in Comparative Perspective*. New York: Palgrave Macmillan
- Stolle, D. (2004): Communities, Social Capital and Local Government: Generalized Trust in Regional Settings. In: Prakash, S.; Selle, P. (Hrsg.): *Investigating Social Capital*. New Delhi: Sage India
- Stolle, D.; Hooghe, M. (2003): Conflicting Approaches to the Study of Social Capital. Competing Explanations for Causes and Effects of Social Capital. *Ethical Perspectives*, 10(1), S. 22-45
- Stolle, D.; Rochon, T. (1998): Are all Associations alike? Member Diversity, Associational Type, and the Creation of Social Capital. *American Behavioral Scientist*, 42(1), S. 47-65

- Stolle, D.; Rochon, T. (1999): The myth of American exceptionalism. A three-nation comparison of associational membership and social capital. In: van Deth, J.; Maraffi, M.; Whiteley, P. (Hrsg.): *Social Capital and European Democracy*. London, Routledge
- Stolle, D.; Rochon, T. (2001): Are all Associations alike? Member Diversity, Associational Type, and the Creation of Social Capital. In: Edwards, B.; Foley, M.; Diani, M. (Hrsg.): *Beyond Tocqueville. Civil Society and the Social Capital Debate in Comparative Perspective*. Hanover: University Press of New England
- Strack, F. (1992): "Order Effects" in Survey Research: Activation and Information Functions of Preceding Questions. In: Schwarz, N.; Sudman, S. (Hrsg.): *Context Effects in Social and Psychological Research*. New York: Springer-Verlag
- Streiner, D. L. (2003): Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, 80(1), S. 99-103
- Sturgis, P.; Smith, P. (2010): Assessing the Validity of Generalized Trust Questions: What Kind of Trust are we measuring? *International Journal of Public Opinion Research*, 22(1), S. 74-92
- Sudman, S.; Blair, E.; Bradburn, N.; Stocking, C. (1977): Estimates of Threatening Behavior Based on Reports of Friends. *Public Opinion Quarterly*, 41(2), S. 261-264
- Sudman, S.; Bradburn, N.M.; Schwarz, N. (1996): *Thinking about Answers. The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers
- Sztompka, P. (1995): Vertrauen: Die fehlende Ressource in der postkommunistischen Gesellschaft. In: Nedelmann, B. (Hrsg.): *Politische Institutionen im Wandel. KZfSS, Sonderheft 35*, S. 254-276
- Sztompka, P. (1996): Trust and Emerging Democracy: Lessons from Poland. *International Sociology*, 11, S. 37-62
- Sztompka, P. (1998): Trust, Distrust and Two Paradoxes of Democracy. *European Journal of Social Theory*, 1(1), S. 19-32
- Sztompka, P. (1999): *Trust. A sociological theory*. Cambridge, UK; New York, NY: Cambridge University Press
- Tabachnick, B. G.; Fidell, L. S. (2001): *Using Multivariate Statistics*. Boston: Allyn and Bacon
- Taylor, M. (1982): *Community, Anarchy, and Liberty*. New York: Cambridge University Press
- Tourangeau, R. (2004): Experimental Design Considerations for Testing and Evaluating Questionnaires. In: Presser, S.; Rothgeb, J.M.; Couper, M.P.; Lessler, J.T.; Martin, E.; Martin, J.; Singer, E. (Hrsg.): *Methods for testing and evaluating survey questionnaires*. New York: Wiley
- Tyler, T. (2001): Why Do People Rely on Others? Social Identity and the Social Aspects of Trust. In: Cook, K. (Hrsg.): *Trust in Society*. New York: Russell Sage Foundation
- Tyler, T. (2003): Trust within Organizations. *Personnel Review*, 32(5), S. 556-568
- Tyler, T.; Degoe, P. (1996): Trust in Organizational Authorities. The Influence of Motive Attributions on Willingness to Accept Decisions. In: Kramer, R. M.; Tyler, T. R. (Hrsg.): *Trust in Organizations. Frontiers of Theory and Research*. Thousand Oaks: Sage

- Überla, K. (1977): *Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler*. Berlin: Springer
- Uslaner, E.M. (1998): Social Capital, Television and the "Mean World": Trust, Optimism and Civic Participation. *Political Psychology*, 19(3), S. 441-467
- Uslaner, E.M. (2001): Volunteering and social capital. How trust and religion shape civic participation in the United States. In: Dekker, P.; Uslaner, E. (Hrsg.): *Social Capital and Participation in Everyday Life*. London: Routledge
- Uslaner, E.M. (2002): *The moral foundations of trust*. Cambridge: Cambridge University Press
- Uslaner, E.M. (2003): Trust, Democracy and Governance: Can Government Policies Influence Generalized Trust? In: Hooghe, M.; Stolle, D. (Hrsg.): *Generating Social Capital. Civil Society and Institutions in Comparative Perspective*. New York: Palgrave Macmillan
- Uslaner, E.M. (2004): Trust and Social Bonds: Faith in Others and Policy Outcomes Reconsidered. *Political Research Quarterly*, 57, S. 501-507
- Uslaner, E.M. (2005): Trust and corruption. In: Graf Lambsdorff, J.; Taube, M.; Schramm, M. (Hrsg.): *The New Institutional Economics of Corruption*. New York: Routledge
- Uslaner, E.M. (2006): The Civil State: Trust, Polarization, and the Quality of State Government. In: Cohen, J. (Hrsg.): *Public Opinion in State Politics*. Stanford: Stanford University Press
- Uslaner, E.M. (2007): The Generalized Trust Questions in the 2006 ANES Pilot Study. Unveröffentlichtes Manuskript
- Uslaner, E.M. (2008a): Trust as a moral value. In: Castiglione, D.; van Deth, J.; Wolleb, G. (Hrsg.): *The handbook of social capital*. Oxford: Oxford University Press
- Uslaner, E.M. (2008b): The foundations of trust: macro and micro. *Cambridge Journal of Economics*, 32, S. 289-294
- Uslaner, E.M. (2008c): Where you stand depends upon where your Grandparents sat. The Inheritability of generalized Trust. *Public Opinion Quarterly*, 72(4), S. 725-740
- Uslaner, E.M. (2010): Is Eleven Really a Lucky Number? Measuring Trust and the Problem of Clumping. Unveröffentlichtes Manuskript
- Uslaner, E.M. (2012a): *Segregation and Mistrust. Diversity, Isolation and Social Cohesion*. Cambridge University Press
- Uslaner, E.M. (2012b): Measuring generalized trust: in defense of the 'standard' question. In: Lyon, F.; Möllering, G.; Saunders, M. (Hrsg.): *Handbook of research methods on trust*. Northampton: Edward Elgar Publishing
- Uslaner, E.M. (2012c): Generalized Trust Questions. In: Aldrich, J.; McGraw, K. (Hrsg.): *Improving Public Opinion Surveys. Interdisciplinary Innovation and the American National Election Studies*. Princeton: Princeton University Press
- Uslaner, E.M. (2013): Trust as an alternative to risk. *Public Choice*, 157, S. 629-639
- Uslaner, E.M.; Brown, M. (2005): Inequality, Trust, and Civic Engagement. *American Politics Research*, 33(6), S. 868-894

- Van Hoorn, A. (2014): Trust Radius versus Trust Level: Radius of Trust as a Distinct Trust Construct. *American Sociological Review*, 79(6), S. 1256-1259
- Van Hoorn, A. (2015): Individualist-Collectivist Culture and Trust Radius: A Multilevel Approach. *Journal of Cross-Cultural Psychology*, 46(2), S. 269-276
- Velicer, W. F.; Fava, J. L. (1998): Effects of Variable and Subject Sampling on Factor Pattern Recovery. *Psychological Methods*, 3(2), S. 231-251
- Walter-Rogg, M. (2005): Politisches Vertrauen ist gut – Misstrauen ist besser? Ausmaß und Ausstrahlungseffekte des Politiker- und Institutionenvertrauens im vereinigten Deutschland. In: Gabriel, Oscar W.; Falter, Jürgen W.; Rattinger, Hans (Hrsg.): *Wächst zusammen, was zusammengehört?* Wiesbaden: Verlag für Sozialwissenschaften
- Warner, S. (1965): Randomized Response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), S. 63-69
- Webb, E.T.; Campbell, D.T.; Schwartz, R.D.; Sechrest, L.; Grove, J.B. (1981): *Nonreactive Measures in the Social Sciences*. Boston: Houghton Mifflin Company
- Weber, H.; Rammsayer, T. (2012): *Differentielle Psychologie-Persönlichkeitsforschung*. Göttingen: Hogrefe
- Weber, L.; Carter, A. (1998): On Constructing Trust: Temporality, Self-Disclosure, and Perspective-Taking. *International Journal of Sociology and Social Policy*, 18(1), S. 7-26
- Weber, L.; Carter, A. (2003): *The Social Construction of Trust*. New York: Kluwer Academic
- Weber, L.; Carter, A. (1997): On Reconstructing Trust: Time, Intention, and Forgiveness. *Clinical Sociology Review*, 15(1), S. 24-39
- Weber, M. (1980): *Wirtschaft und Gesellschaft*. Tübingen: Mohr Siebeck
- Welch, M.; Sikkink, D.; Loveland, M. (2007): The Radius of Trust: Religion, Social Embeddedness and Trust in Strangers. *Social Forces*, 86(1), S. 23-26
- Welzel, C. (2010): How selfish are Self-Expression Values? A Civicness Test. *Journal of Cross-Cultural Psychology*, 41(2), S. 152-174
- Welzel, C.; Delhey, J. (2015): Generalizing Trust: The Benign Force of Emancipation. *Journal of Cross-Cultural Psychology*, 46(7), S. 875-896
- Whiteley, P.F. (1999): The origins of social capital. In: van Deth, J.; Maraffi, M.; Whiteley, P. (Hrsg.): *Social Capital and European Democracy*. London, Routledge
- Winkler, N.; Kroh, M.; Spiess, M. (2006): Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit. DIW, Discussion Paper 579
- Wirtz, M.A.; Strohmer, J. (2013): *Dorsch-Lexikon der Psychologie*. Bern: Verlag Hans Huber
- Wittkowski, J. (2011): Projektive Verfahren. In: Hornke, L.F.; Amelang, M.; Kersting, M. (Hrsg.): *Enzyklopädie der Psychologie. Themenbereich B: Methodologie und Methoden. Serie II: Psychologische Diagnostik. Band 4: Persönlichkeitsdiagnostik*. Göttingen: Hogrefe
- Wolf, C. (1995): Sozioökonomischer Status und berufliches Prestige: ein kleines Kompendium sozialwissenschaftlicher Skalen auf Basis der beruflichen Stellung und Tätigkeit, in: ZUMA-Nachrichten 37, Jg. 19, Neustadt/Weinstraße: Verlag Pfälzische Post GmbH

- Wollebæk, D.; Selle, P. (2002): Does Participation in Voluntary Associations Contribute to Social Capital? The Impact of Intensity, Scope, and Type. *Nonprofit and Voluntary Sector Quarterly*, 31(1), S. 32-61
- Wollebæk, D.; Selle, P. (2003): Participation and Social Capital Formation: Norway in a Comparative Perspective. *Scandinavian Political Studies*, 26(1), S. 67-91
- Wollebæk, D.; Selle, P. (2004): Passive Membership in Voluntary Organizations: Implications for Civil Society, Integration and Democracy. In: Prakash, S.; Selle, P. (Hrsg.): Investigating Social Capital. Comparative Perspectives on Civil Society, Participation and Governance. New Delhi: Sage Publications
- Wollebæk, D.; Selle, P. (2007): Origins of Social Capital: Socialization and Institutionalization Approaches Compared. *Journal of Civil Society*, 3(1), S. 1-24
- Wollebæk, D.; Strømsnes, K. (2008): Voluntary Associations, Trust, and Civic Engagement: A Multilevel Approach. *Nonprofit and Voluntary Sector Quarterly*, 37(2), S. 249-263
- Wright, T.; Tedeschi, R. (1975): Factor Analysis of the Interpersonal Trust Scale. *Journal of Consulting and Clinical Psychology*, 43(4), S. 470-477
- Wrightsmann, L. (1964): Measurement of Philosophies of Human Nature. *Psychological Reports*, 14, S. 743-751
- Wrightsmann, L. (1974): Assumptions about Human Nature: A Social-Psychological Approach. Belmont: Wadsworth Publishing Company
- Yamagishi, T. (2011): Trust. The evolutionary game of mind and society. Tokyo: Springer
- Yamagishi, T.; Yamagishi, M. (1994): Trust and Commitment in the United States and Japan. *Motivation and Emotion*, 18(2), S. 129-166
- Yamagishi, T.; Kikuchi, M.; Kosugi, M. (1999): Trust, gullibility, and social intelligence. *Asian Journal of Social Psychology*, 2, S. 145-161
- Yuki, M.; Maddux, W. W.; Brewer, M. B.; Takemura, K. (2005): Cross-Cultural Differences in Relationship- and Group-Based Trust. *Personality and Social Psychology Bulletin*, 31, S. 48-62
- Ziegler, M.; Schmidt-Atzert, L.; Bühner, M.; Krumm, S. (2007): Fakability of different measurement methods for achievement motivation: questionnaire, semi-projective and objective. *Psychology Science*, 49(4), S. 291-307
- Ziegler, M.; Schmukle, S.; Egloff, B.; Bühner, M. (2010): Investigating Measures of Achievement Motivation(s). *Journal of Individual Differences*, 31(1), S. 15-21
- Ziegler, R. (1997): Interesse, Vernunft und Moral: zur sozialen Konstruktion von Vertrauen. In: Hradil, S. (Hrsg.): Differenz und Integration. Die Zukunft moderner Gesellschaften. Verhandlungen des 28. Kongresses der Deutschen Gesellschaft für Soziologie in Dresden 1996. Frankfurt: Campus
- Zmerli, S. (2004): Politisches Vertrauen und Unterstützung. In: v. Deth, J. (Hrsg.): Deutschland in Europa. Ergebnisse des European Social Survey 2002-2003. Wiesbaden: Verlag für Sozialwissenschaften
- Zmerli, S.; Newton, K. (2008): Social Trust and attitudes Toward Democracy. *Public Opinion Quarterly*, 72(4), S. 706-724

Zmerli, S.; Newton, K.; Montero, J.R. (2007): Trust in people, confidence in political institutions, and satisfaction with democracy. In: van Deth, J.; Montero, J.R.; Westholm, A. (Hrsg.): Citizenship and Involvement in European Democracies. A comparative Analysis. London: Routledge

Zöfel, P. (2003): Statistik für Psychologen im Klartext. München: Pearson

Zwick, W. R.; Velicer, W. F. (1986): Comparison of Five Rules for Determining the Number of Components to Retain. *Psychological Bulletin*, 99(3), S. 432-442

Anhang

Anhang I: Validität

Interkorrelation der gemittelten Skalensummenwerte von Objektiv- und Standard-Skala

Tab. I.1: Bivariate Korrelationen:

Summenscore Objektiv-Skalen (PR_1 bzw. $PR_{1(KS)}$) und Summenscore Standard-Skala (PT_1)

[Referenzgruppe entspricht *outgroup*] (N=23)

	PT_1^a	PR_1^b	$PR_{1(KS)}^c$
PT_1^a	1	---	---
PR_1^b	,495*	1	---
$PR_{1(KS)}^c$,590**	,709**	1

^a Standard-Skala

^b Objektiv-Skala

^c Objektiv-Kurzskala

* $p < .05$ (zweiseitig), ** $p < .01$ (zweiseitig)

Tab. I.2: Bivariate Korrelationen:

Summenscore Objektiv-Skalen (PT_2 bzw. $PT_{2(KS)}$) und Summenscore Standard-Skala (PR_2)

[Referenzgruppe entspricht *outgroup*] (N=30)

	PR_2^a	PT_2^b	$PT_{2(KS)}^c$
PR_2^a	1	---	---
PT_2^b	,436*	1	---
$PT_{2(KS)}^c$,422*	,622**	1

^a Standard-Skala

^b Objektiv-Skala

^c Objektiv-Kurzskala

* $p < .05$ (zweiseitig), ** $p < .01$ (zweiseitig)

Anhang II: Exkurse Signifikanzprüfungen

II.a Exkurs: Prüfung auf Normalverteilung

II.a1 Exkurs: Kolmogorov-Smirnov-Anpassungstest

Für geringe Fallzahlen bietet sich hierzu der Kolmogorov-Smirnov-Anpassungstest an, der sich auch bei kleinen Stichproben als robust erwiesen hat. (Eid/Gollwitzer/Schmitt (2015), S. 324; Zöfel (2002), S. 112) Dabei wird untersucht, ob ein kontinuierliches Merkmal X bzw. eine metrische Variable in der Population von einer bestimmten theoretisch erwarteten Verteilung (hier: Normalverteilung) abweicht. Die Nullhypothese besagt dabei, dass die zu prüfende Verteilung ($\Phi(x)$) bei keiner Ausprägung der Variablen von der erwarteten Verteilungsform ($\Phi_0(x)$) abweicht, und ist formal wie folgt definiert (Eid/Gollwitzer/Schmitt (2015), S. 322):

$$H_0: \Phi(x) = \Phi_0(x) \text{ für alle } x$$

Die Alternativhypothese behauptet dagegen, dass die Verteilung ($\Phi(x)$) von der erwarteten Verteilung ($\Phi_0(x)$) abweicht und in Wahrheit einer anderen Verteilungsform entspricht. Mindestens eine Ausprägung der Variablen weicht somit von der erwarteten Verteilungsform (hier: Normalverteilung) ab. Formal lautet die Alternativhypothese wie folgt (Eid/Gollwitzer/Schmitt (2015), S. 322):

$$H_1: \Phi(x) \neq \Phi_0(x) \text{ für mindestens ein } x$$

Es folgt daraus, dass die betrachtete Variable dann normalverteilt ist und entsprechend für parametrische Signifikanztests geeignet ist, wenn die H_0 beibehalten werden kann und die H_1 abgelehnt wird. Wird der Kolmogorov-Smirnov-Test dagegen signifikant bzw. wird die H_1 beibehalten und die H_0 abgelehnt, so liegt keine Normalverteilung vor und es sollten nonparametrische Verfahren zur Prüfung einer Unterschiedshypothese verwendet werden. (Eid/Gollwitzer/Schmitt (2015), S. 324; Bortz/Lienert/Boehnke (2008), 81, S. 319f; Zöfel (2002), S. 112)

Als Prüfgröße des Kolmogorov-Smirnov-Anpassungstest wird das Maximum der absoluten Abweichungen bzw. Differenzen D_{\max} zwischen der empirischen und der erwarteten Verteilung verwendet. Liegt der Wert dieser Prüfgröße oberhalb eines kritischen Wertes bzw. entspricht er diesem ($D_{\max} \geq D_{\max}^{\text{krit}}$), so spricht man von einem signifikanten Ergebnis. Der kritische Wert richtet sich dabei nach dem angelegten zweiseitigen Signifikanzniveau α . Sind die der empirischen Verteilung zugrundeliegenden Populationsparameter (Populationsmittelwert und -standardabweichung) unbekannt und müssen auf Basis der empirischen Daten geschätzt werden, so fällt das Testergebnis zu konservativ aus. Das bedeutet, auch starke Abweichungen von der Normalverteilung führen nicht zu einer Ablehnung der Nullhypothese (β -Fehler). Um diesem Effekt zu begegnen entwickelte, LILLIEFORS kritische Werte für D_{\max} , die robuste Testergebnisse ermöglichen, auch wenn die Populationsparameter unbekannt sind. Die LILLIEFORS-Korrekturen bewirken dabei, dass die kritischen Werte kleiner werden, was zur Folge hat, dass Abweichungen von der Normalverteilung eher signifikant werden. (Eid/Gollwitzer/Schmitt (2015), S. 324; Zöfel (2002), S. 112)

Da es sich bei den verschiedenen Experimentalgruppen um jeweils voneinander unabhängige Messungen handelt, wird nachfolgend für jede der Gruppen eine Prüfung auf Normalverteilung durchgeführt.

II.b Exkurs: Parametrische Verfahren

II.b1 Exkurs: T-Test für abhängige Stichproben

Der T-Test für abhängige Stichproben prüft, ob sich die Stichprobenmittelwerte zweier abhängiger Stichproben überzufällig voneinander unterscheiden ($\bar{x}_1 \neq \bar{x}_2$) oder äquivalent ($\bar{x}_1 = \bar{x}_2$) sind. Als Prüfgröße wird dabei die in eine t-verteilte Prüfgröße transformierte Differenz der Stichprobenmittelwerte bzw. die t-transformierte mittlere Differenz der Messwertpaare verwendet ($\bar{x}_d \rightarrow T_{\bar{x}_d}$). Ein signifikanter Unterschied zwischen den eingehenden Stichproben ergibt sich, wenn $T_{\bar{x}_d}$ je nach Richtung des Unterschiedes größer bzw. kleiner als der kritische t-Wert ist. Die Bestimmung dieses kritischen t-Wertes erfolgt dabei anhand des angelegten Signifikanzniveaus und der Freiheitsgrade mit $df=n-1$. (Bortz (2005), S. 143ff; Eid/Gollwitzer/Schmitt (2013), S. 369f; Bühner/Ziegler (2009), S. 242ff)

Mit Hilfe des T-Tests für abhängige Stichproben lassen sich sowohl ungerichtete als auch gerichtete Hypothesen prüfen. Die Nullhypothese für einen ungerichteten T-Test besagt, dass sich die Mittelwerte der den Stichproben zugrundeliegenden Populationen μ_1 und μ_2 nicht voneinander unterscheiden bzw. die mittlere Differenz der Messwertpaare in der Grundgesamtheit null ist (Eid/Gollwitzer/Schmitt (2013), S. 369; Bühner/Ziegler (2009), S. 242f):

$$H_0; \text{ungerichtet: } \mu_1 = \mu_2 \text{ oder } \mu_d = 0$$

Die zugehörige ungerichtete Alternativhypothese macht hingegen die Aussage, dass die zwei Populationsmittelwerte μ_1 und μ_2 voneinander abweichen bzw. die mittlere Differenz der Messwertpaare in der Grundgesamtheit ungleich null ist (Eid/Gollwitzer/Schmitt (2013), S. 369; Bühner/Ziegler (2009), S. 242f):

$$H_1; \text{ungerichtet: } \mu_1 \neq \mu_2 \text{ oder } \mu_d \neq 0$$

Gerichtete Hypothesen geben eine Richtung für den Verlauf der Differenz zwischen den Stichproben vor. Entsprechend behauptet eine gerichtete Nullhypothese, dass einer der Populationsmittelwerte größer bzw. kleiner als der andere Populationsmittelwert ist und dementsprechend auch die mittlere Differenz der Messwertpaare je nach vorgegebener Richtung größer oder kleiner null ist (Eid/Gollwitzer/Schmitt (2013), S. 369; Bühner/Ziegler (2009), S. 242f):

$$H_{0; \text{gerichtet1: } \mu_1 \geq \mu_2 \text{ oder } \mu_d \geq 0$$

$$H_{0; \text{gerichtet2: } \mu_1 \leq \mu_2 \text{ oder } \mu_d \leq 0$$

Eine gerichtete Alternativhypothese stellt dagegen die jeweils entgegengesetzte Behauptung auf. Die Differenz ist in der Grundgesamtheit somit entweder größer oder kleiner null (Eid/Gollwitzer/Schmitt (2013), S. 369; Bühner/Ziegler (2009), S. 242f):

$$H_{1; \text{gerichtet1: } \mu_1 < \mu_2 \text{ oder } \mu_d < 0$$

$$H_{1; \text{gerichtet2: } \mu_1 > \mu_2 \text{ oder } \mu_d > 0$$

II.b2 Exkurs: T-Test für unabhängige Stichproben

Der T-Test für unabhängige Stichproben prüft, ob die Stichprobenmittelwerte zweier unabhängiger Stichproben (\bar{x}_1 und \bar{x}_2) aus zwei Populationen mit den Mittelwerten μ_1 und μ_2 stammen, die identisch

sind ($\mu_1 = \mu_2$) oder überzufällig voneinander abweichen ($\mu_1 \neq \mu_2$). Die t-verteilte Prüfgröße wird dabei anhand der Differenz der Stichprobenmittelwerte ($\bar{x}_1 - \bar{x}_2$) und der auf Basis der Stichprobenvarianzen geschätzten Standardfehler der Differenz ($\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}$) berechnet. Dabei unterscheidet sich die Berechnung je nachdem, ob die Stichproben homogene oder heterogene Varianzen aufweisen. Entsprechendes gilt für die Bestimmung der Freiheitsgrade. Analog zum T-Test für abhängige Stichproben ergibt sich ein signifikantes Ergebnis, wenn die Prüfgröße je nach Richtung des Unterschiedes größer oder kleiner als der kritische t-Wert für das angelegte Signifikanzniveau ist. (Eid/Gollwitzer/Schmitt (2013), S. 331; Bortz (2005), S. 140ff; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 161ff; Bühner/Ziegler (2009), S. 254ff)

Auch für den T-Test für unabhängige Stichproben lassen sich sowohl gerichtete als auch ungerichtete Hypothesen formulieren. Eine ungerichtete Nullhypothese macht die Aussage, dass die Mittelwerte der Verteilungen beider Grundgesamtheiten gleich sind bzw. die Differenz der Populationsmittelwerte gleich null ist (Eid/Gollwitzer/Schmitt (2013), S. 331; Bortz (2005), S. 140; Bühner/Ziegler (2009), S. 255):

$$H_0, \text{ ungerichtet: } \mu_1 = \mu_2 \text{ oder } \mu_1 - \mu_2 = 0$$

Eine ungerichtete Alternativhypothese besagt dagegen, dass die Mittelwerte der Verteilungen in den Populationen unterschiedlich sind bzw. die Differenz der Mittelwerte aus den beiden Grundgesamtheiten ungleich null ist (Eid/Gollwitzer/Schmitt (2013), S. 331; Bortz (2005), S. 140; Bühner/Ziegler (2009), S. 255):

$$H_1, \text{ ungerichtet: } \mu_1 \neq \mu_2 \text{ oder } \mu_1 - \mu_2 \neq 0$$

Eine gerichtete H_0 behauptet, dass entweder der Mittelwert der Verteilung in Grundgesamtheit 1 größer bzw. gleich dem Mittelwert der Verteilung in Grundgesamtheit 2 ist oder umgekehrt (Eid/Gollwitzer/Schmitt (2013), S. 331; Bortz (2005), S. 140; Bühner/Ziegler (2009), S. 255):

$$H_0, \text{ gerichtet1: } \mu_1 \geq \mu_2 \text{ oder } \mu_1 - \mu_2 \geq 0$$

$$H_0, \text{ gerichtet2: } \mu_1 \leq \mu_2 \text{ oder } \mu_1 - \mu_2 \leq 0$$

Während die gerichtete Alternativhypothese entweder aussagt, dass der Mittelwert der Verteilung in Grundgesamtheit 1 größer als der Mittelwert der Verteilung in Grundgesamtheit 2 ist, oder angibt, dass die umgekehrte Relation gültig ist (Eid/Gollwitzer/Schmitt (2013), S. 331; Bortz (2005), S. 140; Bühner/Ziegler (2009), S. 255):

$$H_1, \text{ gerichtet1: } \mu_1 < \mu_2 \text{ oder } \mu_1 - \mu_2 < 0$$

$$H_1, \text{ gerichtet2: } \mu_1 > \mu_2 \text{ oder } \mu_1 - \mu_2 > 0$$

II.b3 Exkurs: Einfaktorielle Varianzanalyse (ANOVA)

Die einfaktorielle Varianzanalyse wird dazu eingesetzt, um den Einfluss einer mehrfach gestuften unabhängigen Variablen auf eine abhängige Variable zu prüfen. Die Varianzanalyse ermöglicht es somit, mehr als zwei Gruppen hinsichtlich ihrer zentralen Tendenz zu vergleichen. (Eid/Gollwitzer/Schmitt (2013), S. 391; Bortz (2005), S. 248ff; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 185ff; Bühner/Ziegler (2009), S. 327ff)

Grundidee der Varianzanalyse ist es, dass sich der Messwert einer Person in insgesamt drei Teile zerlegen lässt (Eid/Gollwitzer/Schmitt (2013), S. 393ff; Bühner/Ziegler (2009), S. 335): (1) das „Grundniveau“ des Verhaltens der Probanden über alle Faktorstufen hinweg (\bar{x}), (2) den Effekt des Treatments auf das Verhalten der Probanden (t_j) und (3) Unterschiede zwischen den Probanden, die nicht auf die Faktorstufen zurückgeführt werden können (e_{mj}). Für den Messwert einer Person m ergibt sich dann je nach Faktorstufe j die folgende Zusammensetzung:

$$\bar{x}_{mj} = \bar{x} + t_j + e_{mj}$$

Im Rahmen der einfaktoriellen Varianzanalyse rekrutiert sich die Gesamtvarianz ($\hat{\sigma}_{tot}^2$) der Messwerte, d.h. die Summe der Abweichungen der Probanden vom Gesamtmittelwert über alle Faktorstufen, demnach aus zwei Quellen: (1) der Treatmentvarianz bzw. der Varianz zwischen den Faktorstufen ($\hat{\sigma}_{treat}^2$) und (2) der Fehlervarianz bzw. der Varianz innerhalb der Faktorstufen ($\hat{\sigma}_{Fehler}^2$). Als Maß für die Variation wird dabei jeweils die Quadratsumme der Abweichungen herangezogen, sodass eine totale Quadratsumme (QS_{tot}), eine Treatmentquadratsumme (QS_{treat}) und eine Fehlerquadratsumme (QS_{Fehler}) resultiert. (Eid/Gollwitzer/Schmid (2013), S. 396ff; Bortz (2005), S. 250ff; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 187f; Backhaus/Erichson/Plinke/Weiber (2008), S. 156) Es gilt dann:

$$QS_{tot} = QS_{treat} + QS_{Fehler}$$

Die Gesamtvarianz ergibt sich dann als Quotient aus der totalen Quadratsumme und den zugehörigen Freiheitsgraden. Gleiches gilt für die Treatmentvarianz und die Fehlervarianz. (Bortz (2005), S. 250ff) Wenn kein Unterschied zwischen den Faktorstufen existiert, fallen die Treatment- und Fehlervarianz gleich aus. Besteht ein Unterschied zwischen den Faktorstufen, so wird dieser durch einen Anstieg der Treatmentvarianz repräsentiert. Mit Hilfe der Varianzanalyse wird somit geprüft, ob die Varianz zwischen den Faktorstufen signifikant größer ist als die Varianz innerhalb der Faktorstufen. Dies geschieht anhand eines F-Tests bzw. einer F-verteilten Prüfgröße. (Bühner/Ziegler (2009), S. 335f; Bortz (2005), S. 255f)

Dabei besagt die Nullhypothese, dass die Populationsmittelwerte der in die Analyse eingehenden Faktorstufen identisch sind bzw. kein Unterschied zwischen den Gruppen hinsichtlich ihrer zentralen Tendenz besteht und sie daher aus der gleichen Grundgesamtheit stammen (Eid/Gollwitzer/Schmitt (2013), S. 399; Bortz (2005), S. 249; Bühner/Ziegler (2009), S. 336; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 187f):

$$H_0: \mu_i = \mu_j \text{ für alle Paare } (i, j), i \neq j$$

Dagegen behauptet die Alternativhypothese, dass sich mindestens zwei der Populationsmittelwerte unterscheiden bzw. sich für mindestens zwei Gruppen ein Unterschied hinsichtlich der zentralen Tendenz ergibt (Eid/Gollwitzer/Schmitt (2013), S. 399; Bortz (2005), S. 249; Bühner/Ziegler (2009), S. 336; Kuckartz/Rädiker/Ebert/Schehl (2013), S. 187f):

$$H_1: \mu_i \neq \mu_j \text{ für mindestens ein Paar } (i, j), i \neq j$$

II.c Exkurs: Nonparametrische Verfahren

II.c1 Test für abhängige Stichproben

Abhängige bzw. verbundene Stichproben bezeichnen im Allgemeinen zwei Stichproben, „deren Objekte jeweils paarweise einander zugeordnet sind“ (Bortz (2005), S. 143) Eine spezielle Form abhängiger Stichproben liegt im Falle einer Messwiederholung an ein und derselben Stichprobe vor. Hierbei werden die Probanden zu zwei Messzeitpunkten mit dem gleichen Messinstrument oder mit zueinander äquivalenten Messinstrumenten untersucht. (Bühner/Ziegler (2009), S. 271)

In der vorliegenden Untersuchung stellen jeweils die Versuchsgruppen, die einem *within-subject*-Design entsprechen, abhängige bzw. verbundene Stichproben bezüglich der verschiedenen Varianten zur Vertrauensmessung dar. Dies ergibt sich daraus, dass alle in diesen Gruppen zusammengefassten Probanden sowohl die Standard-Variante als auch die Objektiv-Variante der Vertrauensmessung bearbeiteten und nachfolgend ein Vergleich dieser zueinander äquivalenten Messinstrumente zur Erfassung generalisierten Vertrauens vollzogen werden soll. In *Gruppe 1: Objektiv/Standard* stellt die Standard-Variante somit eine Messwiederholung der Objektiv-Variante dar, während in *Gruppe 2: Standard/Objektiv* die Objektiv-Variante die Messwiederholung der Standard-Variante darstellt. Die Verbundenheit der Messungen gilt dabei allerdings nur innerhalb der jeweiligen Versuchsgruppe, da es sich bei den Probanden, die in der ersten Versuchsgruppe zusammengefasst wurden, um andere Personen handelt als bei den Probanden, die der zweiten Versuchsgruppe zugeordnet sind. Die Versuchsgruppen 1 und 2 verhalten sich somit zueinander wie unabhängige Stichproben.

II.c1.1 Exkurs: Vorzeichen-Rang-Test von WILCOXON

Ein verteilungsfreies Verfahren zur Prüfung von Unterschieden zwischen zwei Messzeitpunkten innerhalb derselben Stichprobe und die nonparametrische Alternative zum t-Test für abhängige Stichproben stellt der Vorzeichen-Rang-Test von WILCOXON für abhängige Stichproben dar. Der Wilcoxon-Test prüft, ob die Populationsmediane der betrachteten verbundenen Stichproben identisch sind ($\tilde{x}_1 = \tilde{x}_2$) oder sich unterscheiden ($\tilde{x}_1 \neq \tilde{x}_2$). Im Speziellen wird allerdings der Unterschied zwischen zwei Messzeitpunkten anhand der Differenzen zwischen den Messwerten zu Zeitpunkt 1 und Zeitpunkt 2 ($d_m = x_{m2} - x_{m1}$) bzw. den aus den Absolutbeträgen dieser Differenzen ($|d_m|$) resultierenden Rangplätzen betrachtet. Dabei können sich positive und negative Veränderungen zwischen den beiden Messzeitpunkten ergeben, wobei vom Betrag her niedrigere Differenzen einen höheren Rang zugeordnet bekommen. Der nächste Schritt besteht in der Aufsummierung der Rangplätze, für die $d_m > 0$ bzw. $d_m < 0$ gilt, um die Rangsummen bzw. Prüfgrößen W^+ und W^- zu ermitteln. Nulldifferenzen ($d_m = x_{m2} - x_{m1} = 0$) werden nicht in die Analyse miteinbezogen. Unter der Voraussetzung, dass sich die Ränge zufällig auf die positiven und negativen Differenzen verteilen, sollten die positiven und negativen Rangsummen gleich ausfallen, wenn sich die abhängigen Stichproben nicht voneinander unterscheiden. (Bühner/Ziegler (2009), S. 271ff; Sedlmeier/Renkewitz (2013), S. 577ff; Eid/Gollwitzer/Schmitt (2015), S. 379f)

Die Nullhypothese besagt demnach, dass die Mediane der beiden verbundenen Stichproben in der Grundgesamtheit gleich ausfallen bzw. die Rangsummen der beiden Messzeitpunkte in der Population gleich sind. Unterschiede zwischen den Messzeitpunkten sind rein zufällig. Formal bedeutet dies (Bühner/Ziegler (2009), S. 271f; Eid/Gollwitzer/Schmitt (2015), S. 379):

$$H_0; \text{ ungerichtet: } W^+ = W^- \text{ bzw. } \tilde{x}_1 = \tilde{x}_2$$

Die ungerichtete Alternativhypothese behauptet dagegen, dass sich die Mediane bzw. die Rangsummen der beiden Stichproben innerhalb der Population unterscheiden. Mit anderen Worten: Positive Ränge treten häufiger auf als negative Ränge oder umgekehrt (Bühner/Ziegler (2009), S. 271f):

$$H_{1; \text{ungerichtet}}: W^+ \neq W^- \text{ bzw. } \tilde{x}_1 \neq \tilde{x}_2$$

Gerichtete Hypothesen geben dagegen eine Richtung für den angenommenen Unterschied zwischen den beiden Messzeitpunkten an. Entsprechend ergeben sich zwei Möglichkeiten, eine H_0 , und zwei Möglichkeiten, eine H_1 zu formulieren. Die H_0 geht dabei davon aus, dass entweder die positive Rangsumme größer als die negative Rangsumme ist oder umgekehrt:

$$H_{0; \text{gerichtet1}}: W^+ \geq W^-$$

$$H_{0; \text{gerichtet2}}: W^+ \leq W^-$$

Die zugehörigen Alternativhypothesen behaupten jeweils das Gegenteil der aufgestellten Nullhypothesen:

$$H_{1; \text{gerichtet1}}: W^+ < W^- \text{ oder } \tilde{x}_1 > \tilde{x}_2$$

$$H_{1; \text{gerichtet2}}: W^+ > W^- \text{ oder } \tilde{x}_1 < \tilde{x}_2$$

Verbesserungen von Messung 1 zu Messung 2 werden als positiver Wert ausgedrückt und gehen in die positive Rangsumme ein, während Verschlechterungen von Messung 1 zu Messung 2 als negativer Wert ausgedrückt werden und in die negative Rangsumme eingehen. Entsprechend deutet eine größere Häufigkeit positiver Ränge auf eine Verbesserung des Medians vom ersten Messzeitpunkt zum zweiten Messzeitpunkt hin und eine größere Häufigkeit negativer Ränge bedeutet eine Verschlechterung des Medians von Messzeitpunkt 1 zu Messzeitpunkt 2. (Bühner/Ziegler (2009), S. 271ff; Sedlmeier/Renkewitz (2013), S. 578f)

II.c2 Test für unabhängige Stichproben

Voneinander unabhängige Stichproben bezeichnen Stichproben, die sich wechselseitig nicht beeinflussen. (Kuckartz/Rädiker/Ebert/Schehl (2013), S. 160) Die Unabhängigkeit der Stichproben wird dabei dadurch gewährleistet, dass sie sich entweder hinsichtlich bestimmter Merkmale oder hinsichtlich des angewendeten Treatments unterscheiden. (Bühner/Ziegler (2009), S.281)

In der vorliegenden Untersuchung lassen sich die Versuchsgruppen 1 bis 4 wechselseitig als unabhängige Stichproben interpretieren, da sich jeweils die Versuchsbedingungen bzw. das angewendete Treatment sowie die Stichprobenzusammensetzung voneinander unterscheiden. Im Fokus der nachfolgenden Analyse stehen allerdings die entlang des *between-subject*-Designs konzipierten Versuchsgruppen 3 und 4. Die Probanden der *Gruppe 3: Objektiv* bearbeiteten dabei ausschließlich die Objektiv-Variante der Vertrauensmessung, während die Probanden der *Gruppe 4: Standard* ausschließlich die Standard-Variante bearbeiteten.

II.c2.1 Exkurs: Mann-Whitney-U-Test bzw. Wilcoxon-Rangsummen-Test

Ein nonparametrisches Verfahren zur Prüfung des Unterschiedes zwischen zwei unabhängigen Stichproben stellen der Mann-Whitney-U-Test bzw. der Wilcoxon-Rangsummen-Test dar. Beide Tests

sind algebraisch identisch, verwenden allerdings unterschiedliche Prüfgrößen und können als verteilungsfreies Analogon zum t-Test für unabhängige Stichproben angesehen werden. Beide Tests prüfen, ob sich die Mediane der beiden unabhängigen Stichproben signifikant voneinander unterscheiden ($\tilde{x}_1 \neq \tilde{x}_2$). Hierzu werden den Messwerten Rangplätze in aufsteigender Reihenfolge zugewiesen, sodass der geringste Wert den niedrigsten Rangplatz erhält. Im Anschluss wird für jede der Stichproben eine separate Rangsumme (RS_1 und RS_2) berechnet. (Bühner/Ziegler (2009), S. 281ff; Eid/Gollwitzer/Schmitt (2015), S. 343ff; Bortz (2005), S. 150; Rasch/Kubinger (2006), S. 190)

Im Falle des Wilcoxon-Rangsummen-Tests wird die Rangsumme der kleineren Stichprobe (RS_k) bzw. bei gleich großen Stichproben per Konvention die Rangsumme der ersten Stichprobe (RS_1) als Prüfgröße W verwendet. Die Rangsummen RS_1 und RS_2 nehmen nahezu gleiche Werte an, wenn sich die beiden unabhängigen Stichproben nicht voneinander unterscheiden. (Eid/Gollwitzer/Schmitt (2015), S. 344)

Im Zuge des U-Tests werden mit Hilfe der Rangsummen zunächst die Anzahl der Rangplatzüberschreitungen $U_{\bar{0}}$ bzw. U_1 und die Anzahl der Rangplatzunterschreitungen $U_{\bar{1}}$ bzw. U_2 bezüglich der ersten Stichprobe bestimmt. Anders ausgedrückt bezeichnet U_1 die Anzahl der Fälle, in denen der Rangplatz eines Probanden aus Stichprobe 1 von Probanden aus Stichprobe 2 überschritten wird, während U_2 die Anzahl der Fälle bezeichnet, in denen der Rangplatz eines Probanden aus Stichprobe 2 von Probanden aus Stichprobe 1 überschritten wird. Die kleinere der beiden Prüfgrößen U_1 bzw. U_2 wird dann als Prüfgröße U verwendet. (Eid/Gollwitzer/Schmitt (2015), S. 345f; Bühner/Ziegler (2009), S. 283ff; Bortz (2005), S. 150f)

Eine ungerichtete Nullhypothese behauptet demnach, dass der Unterschied zwischen den Medianen der beiden unabhängigen Stichproben gleich null ist bzw. in der Grundgesamtheit gleich viele Rangplatzüberschreitungen wie -unterschreitungen vorliegen. Formal ausgedrückt bedeutet dies (Eid/Gollwitzer/Schmitt (2015), S. 343; Bühner/Ziegler (2009), S. 282):

$$H_{0, \text{ungerichtet}}: \tilde{x}_1 = \tilde{x}_2$$

Entsprechend besagt eine ungerichtete Alternativhypothese, dass die Abweichung zwischen den Medianen der beiden unabhängigen Stichproben ungleich 0 ist bzw. in der Grundgesamtheit ungleich viele Rangplatzüberschreitungen und -unterschreitungen vorliegen (Eid/Gollwitzer/Schmitt (2015), S. 343; Bühner/Ziegler (2009), S. 282):

$$H_{1, \text{ungerichtet}}: \tilde{x}_1 \neq \tilde{x}_2$$

Die gerichteten Pendanten der beiden Hypothesen geben dagegen jeweils eine Richtung für den Unterschied an. So besagt die gerichtete H_0 entweder, dass der Median der einen Stichprobe größer oder gleich dem Median der anderen Stichprobe ist, oder umgekehrt:

$$H_{0, \text{gerichtet1}}: \tilde{x}_1 \geq \tilde{x}_2$$

$$H_{0, \text{gerichtet2}}: \tilde{x}_1 \leq \tilde{x}_2$$

Die gerichtete Alternativhypothese behauptet dabei jeweils das Gegenteil der zugehörigen Nullhypothese. So liegen entweder mehr Rangplatzüberschreitungen als -unterschreitungen in der Grundgesamtheit vor oder es liegen mehr Rangplatzunterschreitungen als -überschreitungen in der Grundgesamtheit vor:

$$H_1; \text{gerichtet1: } \tilde{x}_1 < \tilde{x}_2$$

$$H_1; \text{gerichtet1: } \tilde{x}_1 > \tilde{x}_2$$

Anhang III: Partikulares vs. Generalisiertes Vertrauen

III.a Innerhalb der Experimentalgruppen und Messintention identifiziert (Objektiv-Skalen)

Tab. III.a.1: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Partikulares vs. Generalisiertes Vertrauen [Innerhalb der Experimentalgruppen und Messintention identifiziert] (Objektiv-Skala)

PR ₁	T(10)= ,559 (zweiseitig); n.s.; g=,373; 1-β=,080
PT ₂	T(16,000)= 1,376 (zweiseitig); n.s.; g=,391; 1-β=,133
PT ₃	T(9)= ,084 (zweiseitig); n.s.; g=,052; 1-β=,051

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow \text{kleiner Effekt}$, $d = .50 \rightarrow \text{mittlerer Effekt}$, $d = .80 \rightarrow \text{großer Effekt}$ (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

Tab. III.a.2: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Partikulares vs. Generalisiertes Vertrauen [Innerhalb der Experimentalgruppen und Messintention identifiziert] (Objektiv-Kurzskala)

PR _{1(KS)}	T(10)= -,767 (zweiseitig); n.s.; g=,511; 1-β=,107
PT _{2(KS)}	T(22)= -,441 (zweiseitig); n.s.; g=,254; 1-β=,198
PT _{3(KS)}	T(9)= -,405 (zweiseitig); n.s.; g=,254; 1-β=,065

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow \text{kleiner Effekt}$, $d = .50 \rightarrow \text{mittlerer Effekt}$, $d = .80 \rightarrow \text{großer Effekt}$ (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

Tab. III.a.3: Summenscore: Vertrauen entsprechend Referenzgruppe [Messintention identifiziert; Objektiv-Skalen]

Vertrauensreferenz	Experimentalgruppe	\bar{x}_i^c	σ_i^d	N
ingroup	PR ₁ ^a	3,0000	1,00000	3
	PR _{1(KS)} ^b	2,6667	1,15470	3
outgroup	PR ₁ ^a	2,7778	,44096	9
	PR _{1(KS)} ^b	3,1111	,78174	9
ingroup	PR ₂ ^a	3,0000	,00000	7
	PR _{2(KS)} ^b	3,0000	,57735	7
outgroup	PR ₂ ^a	2,8235	,52859	17
	PR _{2(KS)} ^b	3,1176	,60025	17
ingroup	PT ₃ ^a	3,2857	,48795	7
	PT _{3(KS)} ^b	3,1429	,37796	7
outgroup	PT ₃ ^a	3,2500	,95743	4
	PT _{3(KS)} ^b	3,2500	,50000	4

^a Objektiv-Skala

^b Objektiv-Kurzskala

^c Arithmetisches Mittel

^d Standardabweichung

(Quelle: Eigene Berechnung)

III.b Innerhalb Experimentalgruppen und Messintention nicht identifiziert (Objektiv-Skalen)

Tab. III.b.1: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Partikulares vs. Generalisiertes Vertrauen [Innerhalb der Experimentalgruppen und Messintention nicht identifiziert] (Objektiv-Skala)

PR₁	T(35)= -1,087 (zweiseitig); n.s.; g=,359; 1-β=,185
PT₂	T(24)= -,221 (zweiseitig); n.s.; g=,098; 1-β=,055
PT₃	T(37)= -,218 (zweiseitig); n.s.; g=,076; 1-β=,055

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow \text{kleiner Effekt}$, $d = .50 \rightarrow \text{mittlerer Effekt}$, $d = .80 \rightarrow \text{großer Effekt}$ (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

Tab. III.b.2: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Partikulares vs. Generalisiertes Vertrauen [Innerhalb der Experimentalgruppen und Messintention nicht identifiziert] (Objektiv-Kurzskala)

PR_{1(KS)}	T(35)= -,869 (zweiseitig); n.s.; g=,287; 1-β=,135
PT_{2(KS)}	T(24)= -1,223 (zweiseitig); n.s.; g=,541; 1-β=,217
PT_{3(KS)}	T(37)= ,204 (zweiseitig); n.s.; g=,071; 1-β=,055

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow \text{kleiner Effekt}$, $d = .50 \rightarrow \text{mittlerer Effekt}$, $d = .80 \rightarrow \text{großer Effekt}$ (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

Tab. III.b.3: Summenscore: Vertrauen entsprechend Referenzgruppe [Messintention nicht identifiziert; Objektiv-Skalen]

Vertrauensreferenz	Experimentalgruppe	\bar{x}_i^c	σ_i^d	N
ingroup	PR ₁ ^a	2,8235	,52859	17
	PR _{1(KS)} ^b	2,9412	,65865	17
outgroup	PR ₁ ^a	3,0000	,45883	20
	PR _{1(KS)} ^b	3,1000	,44721	20
ingroup	PR ₂ ^a	3,0000	,57735	7
	PR _{2(KS)} ^b	3,0000	,57735	7
outgroup	PR ₂ ^a	3,0526	,52427	19
	PR _{2(KS)} ^b	3,2632	,45241	19
ingroup	PT ₃ ^a	2,9167	,51493	12
	PT _{3(KS)} ^b	3,0833	,66856	12
outgroup	PT ₃ ^a	2,9630	,64935	27
	PT _{3(KS)} ^b	3,0370	,64935	27

^a Objektiv-Skala

^b Objektiv-Kurzskala

^c Arithmetisches Mittel

^d Standardabweichung

(Quelle: Eigene Berechnung)

III.c Innerhalb Experimentalgruppen (Standard-Skala)

Tab. III.c.1: Parametrische Verfahren (T-Test für unabhängige Stichproben)

Partikulares vs. Generalisiertes Vertrauen [Innerhalb der Experimentalgruppen und Messintention identifiziert] (Standard-Skala)

PT ₁	T(47)= -,753 (zweiseitig); n.s.; g=,217; 1-β=,114
PR ₂	T(48)= 1,364 (zweiseitig); n.s.; g=,414; 1-β=,268
PT ₄	T(50)= -,640 (zweiseitig); n.s.; g=,182; 1-β=,096

T(df): empirischem t-Wert inklusive Freiheitsgrade

(einseitig) bzw. (zweiseitig): Testrichtung

*p<.05 (einseitig), **p<.01 (einseitig)

n.s.: nicht signifikant (p>.05 (zweiseitig))

g: Effektstärke mit $d = .20 \rightarrow \text{kleiner Effekt}$, $d = .50 \rightarrow \text{mittlerer Effekt}$, $d = .80 \rightarrow \text{großer Effekt}$ (Eid/Gollwitzer/Schmitt (2015), S. 373f; Cohen (1988), S. 20ff, S. 66ff; Bühner/Ziegler (2009), S. 187, S. 251f)

1-β: Teststärke

Vgl. Bühner/Ziegler (2009), S. 260

Tab. III.c.2: Summenscore: Vertrauen entsprechend Referenzgruppe [Standard-Skala]

Experimentalgruppe	Vertrauensreferenz	\bar{x}_i^b	σ_i^c	N
PT ₁ ^a	ingroup	3,1905	,74960	21
	outgroup	3,3571	,78004	28
PR ₂ ^a	ingroup	3,5625	,96393	16
	outgroup	3,2353	,69887	34
PT ₄ ^a	ingroup	3,1000	,78807	20
	outgroup	3,2500	,84242	32

^a Standard-Skala

^c Arithmetisches Mittel

^d Standardabweichung

(Quelle: Eigene Berechnung)

III.d Zwischen den Experimentalgruppen (Objektiv-Skalen und Standard-Skala)

Tab. III.d.1: Parametrische Verfahren (ANOVA)

Partikulares vs. Generalisiertes Vertrauen [Zwischen den Experimentalgruppen und Messintention identifiziert] (Objektiv-Skalen)

PR _{1Part} vs. PT _{2Part} vs. PT _{3Part} vs. PR _{1Gen} vs. PT _{2Gen} vs. PT _{3Gen}	F(5; 41)= 1,149; n.s.; f=,374; 1-β=,418
PR _{1(KS)Part} vs. PT _{2(KS)Part} vs. PT _{3(KS)Part} vs. PR _{1(KS)Gen} vs. PT _{2(KS)Gen} vs. PT _{3(KS)Gen}	F(5; 41)= ,355; n.s.; f=,207; 1-β=,143

F(df₁;df₂): empirischer F-Wert inklusive Zähler- und Nennerfreiheitsgrade

n.s.: nicht signifikant (p>.05 (zweiseitig))

f: Effektstärke f = .10 → *kleiner Effekt*; f = .25 → *moderater Effekt*; f = .40 → *großer Effekt*

1-β: Teststärke

Vgl. Eid/Gollwitzer/Schmitt (2015), S. 407

Tab. III.d.2: Parametrische Verfahren (ANOVA)

Partikulares vs. Generalisiertes Vertrauen [Zwischen den Experimentalgruppen und Messintention nicht identifiziert] (Objektiv-Skalen)

PR _{1Part} vs. PT _{2Part} vs. PT _{3Part} vs. PR _{1Gen} vs. PT _{2Gen} vs. PT _{3Gen}	F(5; 96)= ,358; n.s.; f=,135; 1-β=,143
PR _{1(KS)Part} vs. PT _{2(KS)Part} vs. PT _{3(KS)Part} vs. PR _{1(KS)Gen} vs. PT _{2(KS)Gen} vs. PT _{3(KS)Gen}	F(5; 96)= ,634; n.s.; f=,182; 1-β=,235

F(df₁;df₂): empirischer F-Wert inklusive Zähler- und Nennerfreiheitsgrade

n.s.: nicht signifikant (p>.05 (zweiseitig))

f: Effektstärke f = .10 → *kleiner Effekt*; f = .25 → *moderater Effekt*; f = .40 → *großer Effekt*

1-β: Teststärke

Vgl. Eid/Gollwitzer/Schmitt (2015), S. 407

Tab. III.d.3: Parametrische Verfahren (ANOVA)

Partikulares vs. Generalisiertes Vertrauen [Zwischen den Experimentalgruppen] (Standard-Skala)

PT _{1Part} vs. PR _{2Part} vs. PT _{4Part} vs. PT _{1Gen} vs. PR _{2Gen} vs. PT _{4Gen}	F(5; 145)= ,743; n.s.; f=,160; 1-β=,273
--	---

F(df₁;df₂): empirischer F-Wert inklusive Zähler- und Nennerfreiheitsgrade

n.s.: nicht signifikant (p>.05 (zweiseitig))

f: Effektstärke f = .10 → *kleiner Effekt*; f = .25 → *moderater Effekt*; f = .40 → *großer Effekt*

1-β: Teststärke

Vgl. Eid/Gollwitzer/Schmitt (2015), S. 407

1.2) Wenn Sie Ihr Leben im Großen und Ganzen betrachten, als wie zufrieden würden Sie sich zum jetzigen Zeitpunkt beschreiben? Kreuzen Sie bitte das Gesicht an, welches Ihrer Lebenszufriedenheit am nächsten kommt.

Bitte lesen Sie jedes Statement aufmerksam und kreuzen Sie an inwieweit Sie der jeweiligen Aussage zustimmen. Wenn Sie eine Aussage nicht bewerten möchten bzw. nicht sinnvoll bewerten können, dann wählen Sie bitte die Antwortalternative „Weiß nicht“ aus.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)	Weiß nicht (0)
2.1) Die Zukunft erscheint vielversprechend.	0	0	0	0	0	0
2.2) Im Allgemeinen handeln alle Menschen auf der Welt nach den gleichen Werten und Normen.	0	0	0	0	0	0
2.3) Wenn man dem Normalbürger eine Aufgabe gibt, so wird er sie erfolgreich erfüllen, ohne dass er überwacht werden muss.	0	0	0	0	0	0
2.4) Wenn man sich anstrengt, wird man in der Regel seine Lebensziele erreichen.	0	0	0	0	0	0
2.5) Alle Menschen teilen allgemeine Ziele und normative Überzeugungen.	0	0	0	0	0	0
2.6) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen.	0	0	0	0	0	0
2.7) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken.	0	0	0	0	0	0
2.8) Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden.	0	0	0	0	0	0
2.9) Wenn jemand seinen Geldbeutel auf dem Tisch eines Restaurants vergisst, wird er in der Mehrzahl der Fälle nicht zurückgegeben.	0	0	0	0	0	0
2.10) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinander treffen, ist von andauernden Konflikten geprägt.	0	0	0	0	0	0

Bitte kreuzen Sie an, inwieweit Sie der jeweiligen Aussage zustimmen!

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)	Weiß nicht (0)
2.11) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.	0	0	0	0	0	0
2.12) Es ist nicht schwer, Menschen zu finden, die einem beim Umzug helfen würden.	0	0	0	0	0	0
2.13) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.	0	0	0	0	0	0
2.14) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.	0	0	0	0	0	0
2.15) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.	0	0	0	0	0	0
2.16) Hat man in irgendeiner Angelegenheit mit staatlichen Behörden zu tun, so wird man in der Regel fair und gerecht behandelt.	0	0	0	0	0	0
2.17) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte.	0	0	0	0	0	0
2.18) Der Umgang mit Personen aus dem näheren persönlichen Umfeld gestaltet sich einfacher, als der Umgang mit Fremden.	0	0	0	0	0	0
2.19) Es spielt keine Rolle, ob eine Wohngegend als sicher gilt, man sollte seine Wohnungstür immer abschließen, wenn man das Haus verlässt.	0	0	0	0	0	0
2.20) Der Durchschnittsmensch ist weitestgehend Herr über sein eigenes Schicksal.	0	0	0	0	0	0
2.21) Der persönliche Einsatz für Vereine und Organisationen stellt eine erfüllende Freizeitbeschäftigung dar.	0	0	0	0	0	0

Bitte kreuzen Sie an, inwieweit Sie der jeweiligen Aussage zustimmen!

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)	Weiß nicht (0)
2.22) Erst wenn man bekommen hat, was man vom Leben erwartet, sollte man sich mit der Ungerechtigkeit in der Welt beschäftigen.	0	0	0	0	0	0
2.23) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden.	0	0	0	0	0	0
2.24) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist.	0	0	0	0	0	0
2.25) Wir verlieren langsam unsere persönliche Freiheit, weil die Behörden sich immer weiter ausdehnen.	0	0	0	0	0	0
2.26) Auf die Dauer gesehen, gleichen sich die guten und schlechten Dinge im Leben aus.	0	0	0	0	0	0
2.27) Eine Kooperation von Personen, die sich ziemlich ähnlich sind, wird erfolgreicher sein, als eine Kooperation von Personen, die sich stark voneinander unterscheiden.	0	0	0	0	0	0
2.28) Wenn man möchte, dass die Leute einen Auftrag korrekt ausführen, sollte man Ihnen die Aufgabe detailliert erklären und sie genau beaufsichtigen.	0	0	0	0	0	0
2.29) Man sollte nur dann den wahren Grund seiner Absichten zeigen, wenn es einem nützt.	0	0	0	0	0	0
2.30) Die meisten Menschen sind sich gar nicht bewusst, wie sehr ihr Leben von Zufällen beeinflusst wird.	0	0	0	0	0	0
2.31) Das Engagement in politischen Parteien sollte innerhalb der Gesellschaft zu einem allgemeingültigen Wert werden.	0	0	0	0	0	0
2.32) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.	0	0	0	0	0	0



Bitte kreuzen Sie an, inwieweit Sie der jeweiligen Aussage zustimmen!

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)	Weiß nicht (0)
2.33) Es ist schwer einschätzbar, wie sich das Handeln der Menschen von Situation zu Situation verändert.	0	0	0	0	0	0
2.34) Es ist möglich, in jeder Hinsicht ein guter Mensch zu sein.	0	0	0	0	0	0
2.35) Der Mensch sollte nur diejenigen staatlichen Beschränkungen seiner Freiheit befolgen, die er selber als vernünftig ansieht.	0	0	0	0	0	0
2.36) Man kann ein Versprechen ruhig brechen, wenn es für einen selbst vorteilhaft ist.	0	0	0	0	0	0
2.37) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.	0	0	0	0	0	0
2.38) Die meisten Menschen haben einen guten Eindruck davon, was ihre Stärken und Schwächen sind.	0	0	0	0	0	0
2.39) Wenn man einen Gegenstand an eine andere Person verleiht, wird man diesen intakt zurückbekommen.	0	0	0	0	0	0
2.40) Heutzutage ist es nicht schwer Freunde zu finden.	0	0	0	0	0	0
2.41) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.	0	0	0	0	0	0
2.42) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.	0	0	0	0	0	0

Anhang IV: Fragebogen

3.) Nachfolgend möchte ich gerne erfahren welche Gedanken Sie bei der Beantwortung der vorangegangenen Fragebatterie hatten.

3.1) Was glauben Sie, wurde mit Hilfe der vorangegangenen Statements abgefragt bzw. welches Konstrukt sollte ihrer Meinung mit den vorangegangenen Aussagen gemessen werden?

Bitte fassen Sie Ihre Annahme nach Möglichkeit in einem oder mehreren Schlagworten zusammen!

3.2) Was hat Sie zu Ihrer Antwort auf die vorangegangene Frage bewogen bzw. woran glauben Sie die Intention des Fragebogens erkannt zu haben?

Bitte geben Sie nach Möglichkeit das jeweilige Statement sinngemäß oder wörtlich wieder!

3.3) An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht?

Bitte wählen Sie nur eine der Antwortalternativen aus.

<input type="radio"/>	Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)
<input type="radio"/>	Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)
<input type="radio"/>	Personen, die Sie erst ein einziges Mal getroffen haben.
<input type="radio"/>	Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)
<input type="radio"/>	Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)

Anhang IV: Fragebogen

4.) Nachstehend sind eine Anzahl von Behauptungen aufgeführt, die persönliche Eigenschaften und Einstellungen betreffen.

Lesen Sie bitte jeden Satz und bestimmen Sie, ob die Behauptung in Bezug auf Sie selbst richtig oder falsch ist, und kreuzen Sie jeweils "stimmt" oder "stimmt nicht" an.

	Stimmt nicht (0)	Stimmt (1)
4.1) Ich sage immer, was ich denke.	<input type="radio"/>	<input type="radio"/>
4.2) Ich bin manchmal ärgerlich, wenn ich meinen Willen nicht bekomme.	<input type="radio"/>	<input type="radio"/>
4.3) Ich bin immer gewillt, einen Fehler, den ich mache, auch zuzugeben.	<input type="radio"/>	<input type="radio"/>
4.4) Ich habe gelegentlich mit Absicht etwas gesagt, was die Gefühle des anderen verletzen könnte.	<input type="radio"/>	<input type="radio"/>

5.) Zum Schluss bitte ich Sie noch um einige persönliche Angaben:

5.1) Welches Geschlecht haben Sie? ☐ männlich ☐ weiblich

5.2) In welchem Jahr sind Sie geboren? _____

5.3) Welches Hauptfach studieren Sie? _____

5.4) Im wievielten Fachsemester studieren Sie derzeit: _____

5.5) Im wievielten Studiensemester studieren Sie derzeit: _____

5.6) Welchen Beruf übt Ihr Vater aus? _____

5.7) Welchen Beruf übt Ihre Mutter aus? _____

5.8) Wie viel Geld steht Ihnen durchschnittlich pro Monat (inkl. Miete, weitere Fixkosten etc.) zur Verfügung?

<input type="radio"/>	Bis 400€
<input type="radio"/>	401€ bis 600€
<input type="radio"/>	601€ bis 800€
<input type="radio"/>	801€ bis 1000€
<input type="radio"/>	1001€ bis 1200€
<input type="radio"/>	Mehr als 1200€

5.9) In welchem Bundesland bzw. welcher Provinz haben Sie bisher die meiste Zeit Ihres Lebens verbracht?

5.10) Haben Sie bisher die meiste Zeit Ihres Lebens (mindestens $\frac{1}{4}$ der Lebenszeit) am gleichen Ort (Dorf, Stadt etc.) gelebt oder sind Sie bereits einmal/mehrfach umgezogen?

<input type="radio"/>	Habe die meiste Zeit an einem Ort gelebt.
<input type="radio"/>	Bin bereits einmal/mehrfach innerhalb eines Ortes umgezogen.
<input type="radio"/>	Bin bereits einmal/mehrfach in andere Orte umgezogen.

Anhang IV: Fragebogen

5.11) Ist Ihr soziales Umfeld (Familie, Freunde, Bekannte etc.) die meiste Zeit Ihres Lebens (mindestens $\frac{3}{4}$ der Lebenszeit) in seiner Zusammensetzung gleich geblieben oder gab es viele Veränderungen?

<input type="radio"/>	Mein soziales Umfeld ist in seiner Zusammensetzung gleich geblieben.
<input type="radio"/>	Mein soziales Umfeld ist in seiner Zusammensetzung eher gleich geblieben.
<input type="radio"/>	Mein soziales Umfeld hat sich in seiner Zusammensetzung eher verändert.
<input type="radio"/>	Mein soziales Umfeld hat sich in seiner Zusammensetzung stark verändert.

6.) Möchten Sie zu dieser Befragung oder zum besseren Verständnis Ihrer Antworten noch etwas anmerken?






Ist Ihnen während der Teilnahme an dieser Befragung etwas negativ aufgefallen? Waren die Fragen an einer Stelle nicht klar oder war Ihnen die Beantwortung unangenehm? Bitte schreiben Sie kurz ein paar Stichworte dazu.

Gewinnspielteilnahme „Befragung zum gesellschaftlichen Miteinander“

Name, Vorname:

Email-Adresse:

Befragung zur Perzeption anderer Menschen (Teil I)

Beurteilen Sie: Wie zufrieden sind Sie mit dem Leben, das Sie jetzt führen?					
 (1)	 (2)	 (3)	 (4)	 (5)	Keine Angabe
0	0	0	0	0	0

Anhang IV: Fragebogen

2.) Auf den folgenden vier Seiten finden Sie eine Serie von Einstellungsstatements. Jedes repräsentiert eine allgemein verbreitete/landläufige Überzeugung, wobei keine per se als richtig oder falsch eingestuft werden kann. Sie werden voraussichtlich mit einigen der Aussagen übereinstimmen und anderen widersprechen. Ich bin daran interessiert in welchem Ausmaß Sie mit den hier vorgebrachten Aussagen übereinstimmen bzw. im Widerspruch stehen.

Bitte lesen Sie jedes Statement aufmerksam und kreuzen Sie an inwieweit Sie der jeweiligen Aussage zustimmen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
2.1) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.2) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.3) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinander treffen, ist von andauernden Konflikten geprägt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.4) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.5) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.6) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.7) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.8) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.9) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.10) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.11) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.12) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.13) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.14) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anhang IV: Fragebogen

3.) Nachfolgend möchte ich gerne erfahren welche Gedanken Sie bei der Beantwortung der vorangegangenen Fragebatterie hatten.

3.1) Was glauben Sie, wurde mit Hilfe der vorangegangenen Statements abgefragt bzw. welches Konstrukt sollte ihrer Meinung mit den vorangegangenen Aussagen gemessen werden?

Bitte fassen Sie Ihre Annahme nach Möglichkeit in einem oder mehreren Schlagworten zusammen!

3.2) Was hat Sie zu Ihrer Antwort auf die vorangegangene Frage bewogen bzw. woran glauben Sie die Intention des Fragebogens erkannt zu haben?

Bitte geben Sie die Nummer des jeweilige Statements an.

3.3) An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht?

Bitte wählen Sie nur eine der Antwortalternativen aus.

<input type="radio"/>	Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)
<input type="radio"/>	Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)
<input type="radio"/>	Personen, die Sie erst ein einziges Mal getroffen haben.
<input type="radio"/>	Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)
<input type="radio"/>	Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)

Anhang IV: Fragebogen

4.) Nachfolgend sind eine Reihe von Behauptungen aufgeführt, die menschliche Verhaltensweisen betreffen.Lesen Sie bitte jeden Satz und bestimmen Sie, inwieweit die Behauptung in Bezug auf Sie selbst zu trifft.

	Trifft überhaupt nicht zu (1)	Trifft nicht zu (2)	Trifft weniger zu (3)	Trifft teils zu, teils nicht zu (4)	Trifft eher zu (5)	Trifft zu (6)	Trifft voll und ganz zu (7)
4.1) Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.2) Ich bin mir oft unsicher in meinem Urteil.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.3) Ich weiß immer genau, wieso ich etwas mag.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.4) Ich habe schon mal zu viel Wechselgeld zurückbekommen und nichts gesagt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.5) Ich bin immer ehrlich zu anderen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.6) Ich habe gelegentlich mal jemanden ausgenutzt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5.) Welche Einstellung haben Sie zu politischen Institutionen und im Speziellen zu sogenannte Implementationsinstitutionen, wie Judikative und Administration?

Bitte geben Sie inwieweit Sie der jeweiligen Aussage zustimmend oder ablehnend gegenüber stehen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
5.1) Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.2) Hat man in irgendeiner Angelegenheit mit staatlichen Behörden zu tun, so wird man in der Regel fair und gerecht behandelt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.3) Wir verlieren langsam unsere persönliche Freiheit, weil die Behörden sich immer weiter ausdehnen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.4) Die öffentliche Sicherheit ist durch die Behörden gewährleistet.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6.) Die nachfolgenden Statements betreffen Ihre Eingebundenheit in informelle soziale Netzwerke.

Bitte kreuzen Sie für jede der folgenden Aussagen an, ob Sie ihr eher zustimmen oder nicht zustimmen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
6.1) Es ist nicht schwer, Menschen zu finden, die einem beim Umzug helfen würden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.2) Einsamkeit stellt heutzutage ein allgegenwärtiges Gefühl dar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.3) Es gibt nur wenige Menschen, mit denen man wirklich reden kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.4) Heutzutage ist es nicht schwer Freunde zu finden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anhang IV: Fragebogen

7.) Inwieweit sind Sie in formellen Netzwerken engagiert?

Bitte geben Sie an, ob Sie Mitglied in einem der folgenden Typen von Vereinen und Organisationen sind.
(Mehrfachnennungen sind möglich!!!)

<input type="radio"/>	Gewerkschaften
<input type="radio"/>	Partei
<input type="radio"/>	Bürgerinitiative/Nachbarschaftsinitiative
<input type="radio"/>	Umweltschutzgruppe
<input type="radio"/>	Wohltätigkeitsverein
<input type="radio"/>	Kirchliche Gruppe
<input type="radio"/>	Kultureller Verein
<input type="radio"/>	Sportverein/Freizeitverein
<input type="radio"/>	Andere:
<input type="radio"/>	Keine Mitgliedschaft

Bitte fahren Sie fort mit Teil II der Befragung.

IV.b2 Hauptstudie (*Experimentalgruppe 1: Objektiv/Standard*)

Standard-Variante

Befragung zur Perzeption anderer Menschen (Teil II)

8.) Als nächstes würde ich gerne erfahren wie groß das Vertrauen ist, das Sie anderen Menschen entgegenbringen.

Bitte kreuzen Sie für jede der nachfolgenden Aussagen an, ob eher die linke oder die rechte Seite der jeweiligen Antwortskala Ihre Meinung widerspiegelt.

8.1) Ganz allgemein gesprochen: Glauben Sie, dass man den meisten Menschen vertrauen kann, oder dass man im Umgang mit anderen Menschen nicht vorsichtig genug sein kann?

<i>Man kann vertrauen.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Man kann nicht vorsichtig genug sein.</i>
----------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	--

8.2) Und glauben Sie, dass die Menschen versuchen hilfsbereit zu sein, oder dass die Menschen meistens auf den eigenen Vorteil bedacht sind?"

<i>Versuchen hilfsbereit zu sein.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Sind auf den eigenen Vorteil bedacht.</i>
---------------------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	--

8.3) Glauben Sie, dass die meisten Menschen versuchen Sie auszunutzen, wenn Sie die Gelegenheit dazu haben, oder versuchen die meisten Menschen, sich fair zu verhalten?

<i>Versuchen einen auszunutzen.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Versuchen sich fair zu verhalten.</i>
-------------------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	--

9.) An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht?

Bitte wählen Sie nur eine der Antwortalternativen aus.

<input type="radio"/>	Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)
<input type="radio"/>	Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)
<input type="radio"/>	Personen, die Sie erst ein einziges Mal getroffen haben.
<input type="radio"/>	Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)
<input type="radio"/>	Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)

10.) Zum Schluss bitte ich Sie noch um einige persönliche Angaben:

10.1) Welches Geschlecht haben Sie? ☐ männlich ☐ weiblich

10.2) In welchem Jahr sind Sie geboren? _____

10.3) Welches Fach studieren Sie? _____

Vielen Dank für Ihre Teilnahme!!!



Gewinnspielteilnahme „Befragung zum Vertrauen in andere Menschen“

Name, Vorname:






Email-Adresse:

Befragung zur Perzeption anderer Menschen (Teil I)



Universität Regensburg

Vielen Dank für Ihre Mithilfe!

					Keine Angabe
(1)	(2)	(3)	(4)	(5)	
0	0	0	0	0	0

Anhang IV: Fragebogen

2.) Als nächstes würde ich gerne erfahren wie groß das Vertrauen ist, das Sie anderen Menschen entgegenbringen.

Bitte kreuzen Sie für jede der nachfolgenden Aussagen an, ob eher die linke oder die rechte Seite der jeweiligen Antwortskala Ihre Meinung widerspiegelt.

2.1) Ganz allgemein gesprochen: Glauben Sie, dass man den meisten Menschen vertrauen kann, oder dass man im Umgang mit anderen Menschen nicht vorsichtig genug sein kann?

Man kann vertrauen.	0	0	0	0	0	Man kann nicht vorsichtig genug sein.
---------------------	---	---	---	---	---	---------------------------------------

2.2) Und glauben Sie, dass die Menschen versuchen hilfsbereit zu sein, oder dass die Menschen meistens auf den eigenen Vorteil bedacht sind?“

Versuchen hilfsbereit zu sein.	0	0	0	0	0	Sind auf den eigenen Vorteil bedacht.
--------------------------------	---	---	---	---	---	---------------------------------------

2.3) Glauben Sie, dass die meisten Menschen versuchen Sie auszunutzen, wenn Sie die Gelegenheit dazu haben, oder versuchen die meisten Menschen, sich fair zu verhalten?

Versuchen einen auszunutzen.	0	0	0	0	0	Versuchen sich fair zu verhalten.
------------------------------	---	---	---	---	---	-----------------------------------

3.) An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht?

Bitte wählen Sie nur eine der Antwortalternativen aus.

Bitte wählen Sie nur <u>eine</u> der Antwortalternativen aus!	
<input type="radio"/>	Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)
<input type="radio"/>	Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)
<input type="radio"/>	Personen, die Sie erst ein einziges Mal getroffen haben.
<input type="radio"/>	Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)
<input type="radio"/>	Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)

4.) Nachfolgend sind eine Reihe von Behauptungen aufgeführt, die menschliche Verhaltensweisen betreffen.

Lesen Sie bitte jeden Satz und bestimmen Sie, inwieweit die Behauptung in Bezug auf Sie selbst zu trifft.

	Trifft überhaupt nicht zu (1)	Trifft nicht zu (2)	Trifft weniger zu (3)	Trifft teils zu, teils nicht zu (4)	Trifft eher zu (5)	Trifft zu (6)	Trifft voll und ganz zu (7)
4.1) Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus.	O	O	O	O	O	O	O
4.2) Ich bin mir oft unsicher in meinem Urteil.	O	O	O	O	O	O	O
4.3) Ich weiß immer genau, wieso ich etwas mag.	O	O	O	O	O	O	O
4.4) Ich habe schon mal zu viel Wechselgeld zurückbekommen und nichts gesagt.	O	O	O	O	O	O	O
4.5) Ich bin immer ehrlich zu anderen.	O	O	O	O	O	O	O
4.6) Ich habe gelegentlich mal jemanden ausgenutzt.	O	O	O	O	O	O	O

Anhang IV: Fragebogen

5.) Welche Einstellung haben Sie zu politischen Institutionen und im Speziellen zu sogenannte Implementationsinstitutionen, wie Judikative und Administration?

Bitte geben Sie an inwieweit Sie der jeweiligen Aussage zustimmend oder ablehnend gegenüber stehen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
5.1) Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.2) Hat man in irgendeiner Angelegenheit mit staatlichen Behörden zu tun, so wird man in der Regel fair und gerecht behandelt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.3) Wir verlieren langsam unsere persönliche Freiheit, weil die Behörden sich immer weiter ausdehnen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.4) Die öffentliche Sicherheit ist durch die Behörden gewährleistet.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6.) Die nachfolgenden Statements betreffen Ihre Eingebundenheit in informelle soziale Netzwerke.

Bitte kreuzen Sie für jede der folgenden Aussagen an, ob Sie ihr eher zustimmen oder nicht zustimmen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
6.1) Es ist nicht schwer, Menschen zu finden, die einem beim Umzug helfen würden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.2) Einsamkeit stellt heutzutage ein allgegenwärtiges Gefühl dar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.3) Es gibt nur wenige Menschen, mit denen man wirklich reden kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.4) Heutzutage ist es nicht schwer Freunde zu finden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7.) Inwieweit sind Sie in formellen Netzwerken engagiert?

Bitte geben Sie an, ob Sie Mitglied in einem der folgenden Typen von Vereinen und Organisationen sind.

(Mehrfachnennungen sind möglich!!!)

<input type="radio"/>	Gewerkschaften
<input type="radio"/>	Partei
<input type="radio"/>	Bürgerinitiative/Nachbarschaftsinitiative
<input type="radio"/>	Umweltschutzgruppe
<input type="radio"/>	Wohltätigkeitsverein
<input type="radio"/>	Kirchliche Gruppe
<input type="radio"/>	Kultureller Verein
<input type="radio"/>	Sportverein/Freizeitverein
<input type="radio"/>	Andere:
<input type="radio"/>	Keine Mitgliedschaft

Bitte fahren Sie fort mit Teil II der Befragung.

IV.c2 Hauptstudie (Experimentalgruppe 2: Standard/Objektiv)**Objektiv-Variante****Befragung zur Perzeption anderer Menschen (Teil II)**

8.) Auf der folgenden Seite finden Sie eine Serie von Einstellungsstatements. Jedes repräsentiert eine allgemein verbreitete/landläufige Überzeugung, wobei keine per se als richtig oder falsch eingestuft werden kann. Sie werden voraussichtlich mit einigen der Aussagen übereinstimmen und anderen widersprechen. Ich bin daran interessiert in welchem Ausmaß Sie mit den hier vorgebrachten Aussagen übereinstimmen bzw. im Widerspruch stehen.

Bitte lesen Sie jedes Statement aufmerksam und kreuzen Sie an inwieweit Sie der jeweiligen Aussage zustimmen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
8.1) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.2) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.3) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinander treffen, ist von andauernden Konflikten geprägt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.4) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.5) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.6) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.7) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.8) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.9) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.10) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.11) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.12) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.13) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.14) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anhang IV: Fragebogen

9.) Nachfolgend möchte ich gerne erfahren, welche Gedanken Sie bei der Beantwortung der vorangegangenen Fragebatterie hatten.

9.1) Was glauben Sie, wurde mit Hilfe der vorangegangenen Statements abgefragt bzw. welches Konstrukt sollte ihrer Meinung mit den vorangegangenen Aussagen gemessen werden?

Bitte fassen Sie Ihre Annahme nach Möglichkeit in einem oder mehreren Schlagworten zusammen!

9.2) Was hat Sie zu Ihrer Antwort auf die vorangegangene Frage bewogen bzw. woran glauben Sie die Intention des Fragebogens erkannt zu haben?

Bitte geben Sie die Nummer des jeweiligen Statements an.

9.3) An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht?
Bitte wählen Sie nur eine der Antwortalternativen aus.

<input type="radio"/>	Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)
<input type="radio"/>	Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)
<input type="radio"/>	Personen, die Sie erst ein einziges Mal getroffen haben.
<input type="radio"/>	Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)
<input type="radio"/>	Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)

10.) Zum Schluss bitte ich Sie noch um einige persönliche Angaben:

10.1) Welches Geschlecht haben Sie? ☐ männlich ☐ weiblich

10.2) In welchem Jahr sind Sie geboren? _____

10.3) Welches Fach studieren Sie? _____






Vielen Dank für Ihre Teilnahme!!!






Befragung zur Perzeption anderer Menschen



Universität Regensburg

Vielen Dank für Ihre Mithilfe!

					Keine Angabe
(1)	(2)	(3)	(4)	(5)	
0	0	0	0	0	0

					Keine Angabe
(1)	(2)	(3)	(4)	(5)	
0	0	0	0	0	0

Anhang IV: Fragebogen

2.) Auf den folgenden vier Seiten finden Sie eine Serie von Einstellungsstatements. Jedes repräsentiert eine allgemein verbreitete/landläufige Überzeugung, wobei keine per se als richtig oder falsch eingestuft werden kann. Sie werden voraussichtlich mit einigen der Aussagen übereinstimmen und anderen widersprechen. Ich bin daran interessiert in welchem Ausmaß Sie mit den hier vorgebrachten Aussagen übereinstimmen bzw. im Widerspruch stehen.

Bitte lesen Sie jedes Statement aufmerksam und kreuzen Sie an inwieweit Sie der jeweiligen Aussage zustimmen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
2.1) Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.2) Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.3) Das Leben in einer Umgebung, in der verschiedene gesellschaftliche Gruppen (z.B. unterschiedl. Religionen, Nationalitäten oder Ethnien) aufeinander treffen, ist von andauernden Konflikten geprägt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.4) Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.5) Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.6) Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.7) Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.8) Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.9) Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.10) Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.11) Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.12) Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.13) Es ist sicherer Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.14) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anhang IV: Fragebogen

3.) Nachfolgend möchte ich gerne erfahren welche Gedanken Sie bei der Beantwortung der vorangegangenen Fragebatterie hatten.

3.1) Was glauben Sie, wurde mit Hilfe der vorangegangenen Statements abgefragt bzw. welches Konstrukt sollte ihrer Meinung mit den vorangegangenen Aussagen gemessen werden?

Bitte fassen Sie Ihre Annahme nach Möglichkeit in einem oder mehreren Schlagworten zusammen!

3.2) Was hat Sie zu Ihrer Antwort auf die vorangegangene Frage bewogen bzw. woran glauben Sie die Intention des Fragebogens erkannt zu haben?

Bitte geben Sie die Nummer des jeweilige Statements an.

3.3) An welche Personen bzw. Personengruppe haben Sie bei der Beantwortung der Statements hauptsächlich gedacht?

Bitte wählen Sie nur eine der Antwortalternativen aus.

<input type="radio"/>	Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)
<input type="radio"/>	Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)
<input type="radio"/>	Personen, die Sie erst ein einziges Mal getroffen haben.
<input type="radio"/>	Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)
<input type="radio"/>	Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)

Anhang IV: Fragebogen

4.) Nachfolgend sind eine Reihe von Behauptungen aufgeführt, die menschliche Verhaltensweisen betreffen.Lesen Sie bitte jeden Satz und bestimmen Sie, inwieweit die Behauptung in Bezug auf Sie selbst zu trifft.

	Trifft überhaupt nicht zu (1)	Trifft nicht zu (2)	Trifft weniger zu (3)	Trifft teils zu, teils nicht zu (4)	Trifft eher zu (5)	Trifft zu (6)	Trifft voll und ganz zu (7)
4.1) Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.2) Ich bin mir oft unsicher in meinem Urteil.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.3) Ich weiß immer genau, wieso ich etwas mag.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.4) Ich habe schon mal zu viel Wechselgeld zurückbekommen und nichts gesagt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.5) Ich bin immer ehrlich zu anderen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4.6) Ich habe gelegentlich mal jemanden ausgenutzt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5.) Welche Einstellung haben Sie zu politischen Institutionen und im Speziellen zu sogenannte Implementationsinstitutionen, wie Judikative und Administration?

Bitte geben Sie inwieweit Sie der jeweiligen Aussage zustimmend oder ablehnend gegenüber stehen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
5.1) Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.2) Hat man in irgendeiner Angelegenheit mit staatlichen Behörden zu tun, so wird man in der Regel fair und gerecht behandelt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.3) Wir verlieren langsam unsere persönliche Freiheit, weil die Behörden sich immer weiter ausdehnen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.4) Die öffentliche Sicherheit ist durch die Behörden gewährleistet.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6.) Die nachfolgenden Statements betreffen Ihre Eingebundenheit in informelle soziale Netzwerke.

Bitte kreuzen Sie für jede der folgenden Aussagen an, ob Sie ihr eher zustimmen oder nicht zustimmen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
6.1) Es ist nicht schwer, Menschen zu finden, die einem beim Umzug helfen würden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.2) Einsamkeit stellt heutzutage ein allgegenwärtiges Gefühl dar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.3) Es gibt nur wenige Menschen, mit denen man wirklich reden kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.4) Heutzutage ist es nicht schwer Freunde zu finden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anhang IV: Fragebogen

7.) Inwieweit sind Sie in formellen Netzwerken engagiert?

Bitte geben Sie an, ob Sie Mitglied in einem der folgenden Typen von Vereinen und Organisationen sind.

(Mehrfachnennungen sind möglich!!!)

<input type="radio"/>	Gewerkschaften
<input type="radio"/>	Partei
<input type="radio"/>	Bürgerinitiative/Nachbarschaftsinitiative
<input type="radio"/>	Umweltschutzgruppe
<input type="radio"/>	Wohltätigkeitsverein
<input type="radio"/>	Kirchliche Gruppe
<input type="radio"/>	Kultureller Verein
<input type="radio"/>	Sportverein/Freizeitverein
<input type="radio"/>	Andere:
<input type="radio"/>	Keine Mitgliedschaft

8.) Zum Schluss bitte ich Sie noch um einige persönliche Angaben:

8.1) Welches Geschlecht haben Sie? ☐ männlich ☐ weiblich

8.2) In welchem Jahr sind Sie geboren? _____

8.3) Welches Fach studieren Sie? _____

Vielen Dank für Ihre Teilnahme!!!



.....

Gewinnspielteilnahme „Befragung zur Wahrnehmung anderer Menschen“

Name, Vorname:






Email-Adresse:

Befragung zum Vertrauen in andere Menschen



Universität Regensburg

Vielen Dank für Ihre Mithilfe!

					Keine Angabe
(1)	(2)	(3)	(4)	(5)	
0	0	0	0	0	0

Bitte kreuzen Sie für jede der nachfolgenden Aussagen an, ob eher die linke oder die rechte Seite der jeweiligen Antwortskala Ihre Meinung widerspiegelt.

Man kann vertrauen.	0	0	0	0	0	Man kann nicht vorsichtig genug sein.
---------------------	---	---	---	---	---	---------------------------------------

Versuchen hilfsbereit zu sein.	0	0	0	0	0	Sind auf den eigenen Vorteil bedacht.
--------------------------------	---	---	---	---	---	---------------------------------------

Versuchen einen auszunutzen.	0	0	0	0	0	Versuchen sich fair zu verhalten.
------------------------------	---	---	---	---	---	-----------------------------------

<input type="radio"/>	Personen, zu denen Sie eine persönliche bzw. intime Beziehung haben (z.B. Familie, Freunde, Lebenspartner etc.)
<input type="radio"/>	Personen, die Ihnen persönlich bekannt sind und mit denen Sie regelmäßig interagieren (z.B. Nachbarn, Bekannte, Kollegen etc.)
<input type="radio"/>	Personen, die Sie erst ein einziges Mal getroffen haben.
<input type="radio"/>	Personen, die anderen gesellschaftlichen Gruppen angehören als Sie selbst (z.B. andere Religionen, andere Nationalitäten etc.)
<input type="radio"/>	Menschen im Allgemeinen (z.B. Fremde, die Ihnen zufällig auf der Straße begegnen)

Lesen Sie bitte jeden Satz und bestimmen Sie, inwieweit die Behauptung in Bezug auf Sie selbst zu trifft.

	Trifft überhaupt nicht zu (1)	Trifft nicht zu (2)	Trifft weniger zu (3)	Trifft teils zu, teils nicht zu (4)	Trifft eher zu (5)	Trifft zu (6)	Trifft voll und ganz zu (7)
4.1) Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus.	O	O	O	O	O	O	O
4.2) Ich bin mir oft unsicher in meinem Urteil.	O	O	O	O	O	O	O
4.3) Ich weiß immer genau, wieso ich etwas mag.	O	O	O	O	O	O	O
4.4) Ich habe schon mal zu viel Wechselgeld zurückbekommen und nichts gesagt.	O	O	O	O	O	O	O
4.5) Ich bin immer ehrlich zu anderen.	O	O	O	O	O	O	O
4.6) Ich habe gelegentlich mal jemanden ausgenutzt.	O	O	O	O	O	O	O

Anhang IV: Fragebogen

5.) Welche Einstellung haben Sie zu politischen Institutionen und im Speziellen zu sogenannte Implementationsinstitutionen, wie Judikative und Administration?

Bitte geben Sie inwieweit Sie der jeweiligen Aussage zustimmend oder ablehnend gegenüber stehen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
5.1) Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.2) Hat man in irgendeiner Angelegenheit mit staatlichen Behörden zu tun, so wird man in der Regel fair und gerecht behandelt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.3) Wir verlieren langsam unsere persönliche Freiheit, weil die Behörden sich immer weiter ausdehnen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.4) Die öffentliche Sicherheit ist durch die Behörden gewährleistet.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6.) Die nachfolgenden Statements betreffen Ihre Eingebundenheit in informelle soziale Netzwerke.

Bitte kreuzen Sie für jede der folgenden Aussagen an, ob Sie ihr eher zustimmen oder nicht zustimmen.

	Stimme nicht zu (1)	Stimme eher nicht zu (2)	Stimme teils zu, stimme teils nicht zu (3)	Stimme eher zu (4)	Stimme zu (5)
6.1) Es ist nicht schwer, Menschen zu finden, die einem beim Umzug helfen würden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.2) Einsamkeit stellt heutzutage ein allgegenwärtiges Gefühl dar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.3) Es gibt nur wenige Menschen, mit denen man wirklich reden kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6.4) Heutzutage ist es nicht schwer Freunde zu finden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7.) Inwieweit sind Sie in formellen Netzwerken engagiert?

Bitte geben Sie an, ob Sie Mitglied in einem der folgenden Typen von Vereinen und Organisationen sind.

(Mehrfachnennungen sind möglich!!!)

<input type="radio"/>	Gewerkschaften
<input type="radio"/>	Partei
<input type="radio"/>	Bürgerinitiative/Nachbarschaftsinitiative
<input type="radio"/>	Umweltschutzgruppe
<input type="radio"/>	Wohltätigkeitsverein
<input type="radio"/>	Kirchliche Gruppe
<input type="radio"/>	Kultureller Verein
<input type="radio"/>	Sportverein/Freizeitverein
<input type="radio"/>	Andere:
<input type="radio"/>	Keine Mitgliedschaft

8.) Zum Schluss bitte ich Sie noch um einige persönliche Angaben:

- 8.1) Welches Geschlecht haben Sie? ☐ männlich ☐ weiblich
- 8.2) In welchem Jahr sind Sie geboren? _____
- 8.3) Welches Fach studieren Sie? _____

Vielen Dank für Ihre Teilnahme!!!


Gewinnspielteilnahme „Befragung zum Vertrauen in andere Menschen“

Name, Vorname:

Email-Adresse:

Anhang V: Häufigkeitsverteilungen (Pilotstudie)

V.a Rohskala „Generalisiertes Vertrauen“

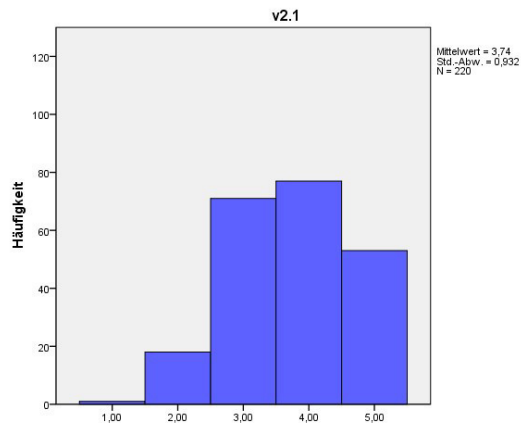


Abb. V.a1: Häufigkeitsverteilung von Item 2.1 „Die Zukunft erscheint vielversprechend.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

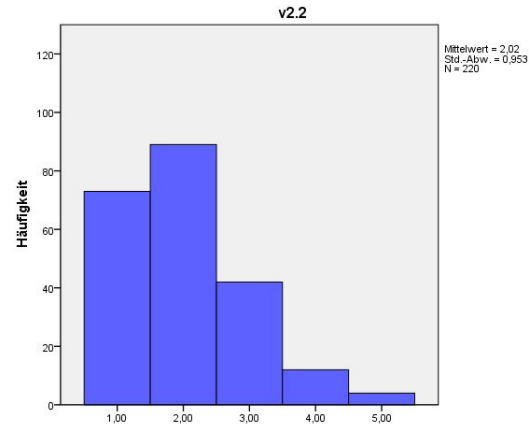


Abb. V.a2: Häufigkeitsverteilung von Item 2.2 „Im Allgemeinen handeln alle Menschen auf der Welt nach den gleichen Werten und Normen.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

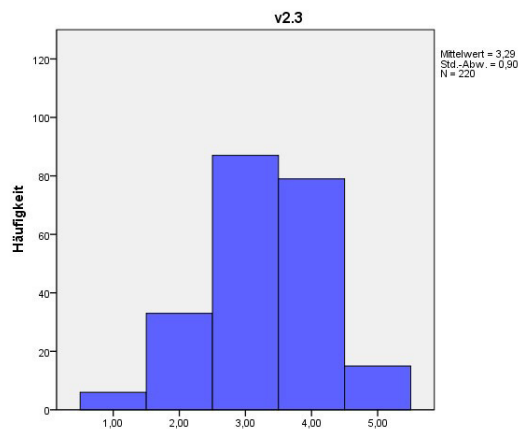


Abb. V.a3: Häufigkeitsverteilung von Item 2.3 „Wenn man dem Normalbürger eine Aufgabe gibt, so wird er sie erfolgreich erfüllen, ohne dass er überwacht werden muss.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

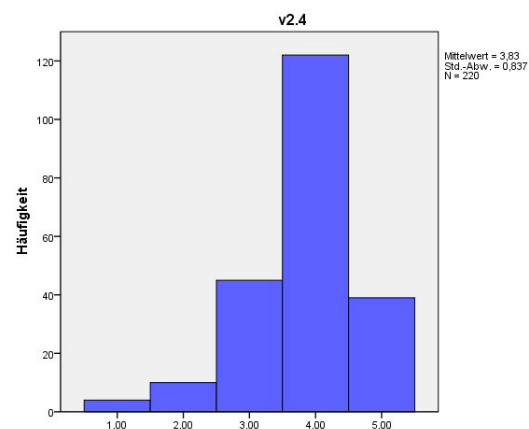


Abb. V.a4: Häufigkeitsverteilung von Item 2.4 „Wenn man sich anstrengt, wird man in der Regel seine Lebensziele erreichen.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

Anhang V: Häufigkeitsverteilungen (Pilotstudie)

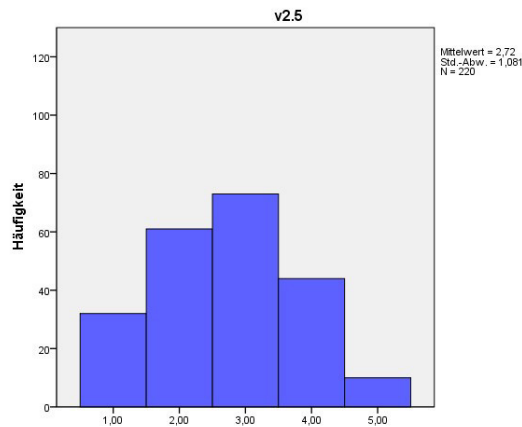


Abb. V.a5: Häufigkeitsverteilung von Item 2.5
„Alle Menschen teilen allgemeine Ziele und normative Überzeugungen.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

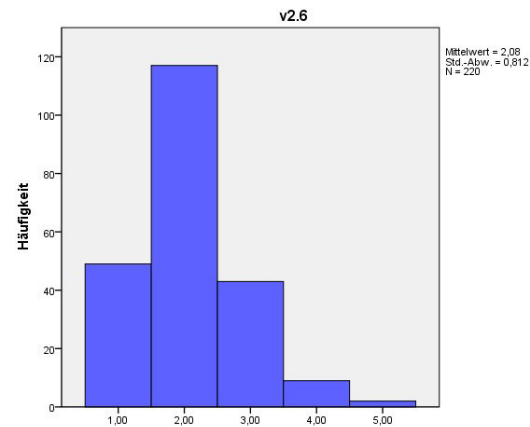


Abb. V.a6: Häufigkeitsverteilung von Item 2.6
„Der Durchschnittsmensch wird eher selten seine Meinung sagen, wenn er bemerkt, dass die anderen ihm nicht zustimmen.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

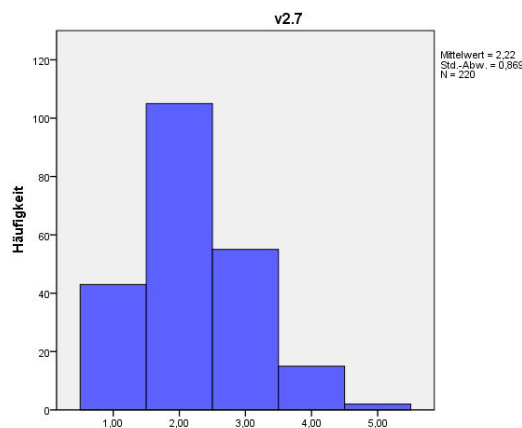


Abb. V.a7: Häufigkeitsverteilung von Item 2.7 „Es ist sicherer zu glauben, dass im Gegensatz zu dem, was die Leute sagen, diese in erster Linie an ihr eigenes Wohlergehen denken.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

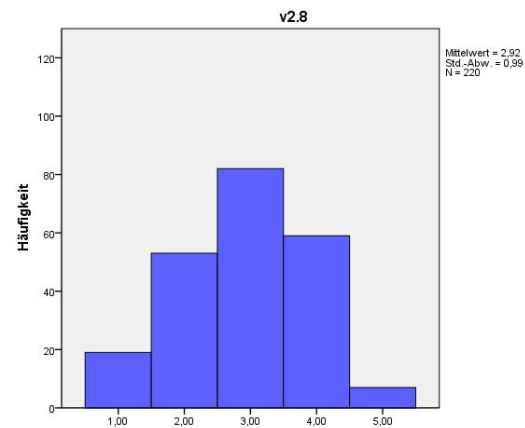


Abb. V.a8: Häufigkeitsverteilung von Item 2.8
„Das Justizsystem stellt einen Bereich dar, in dem alle Menschen unvoreingenommen behandelt werden.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

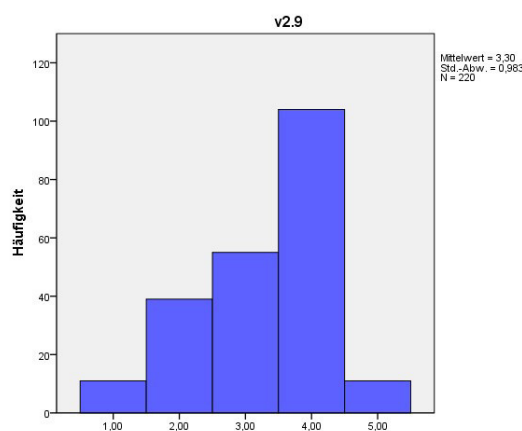


Abb. V.a9: Häufigkeitsverteilung von Item 2.9
„Wenn jemand seinen Geldbeutel auf dem Tisch eines Restaurants vergisst, wird er in der Mehrzahl der Fälle nicht zurückgegeben.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

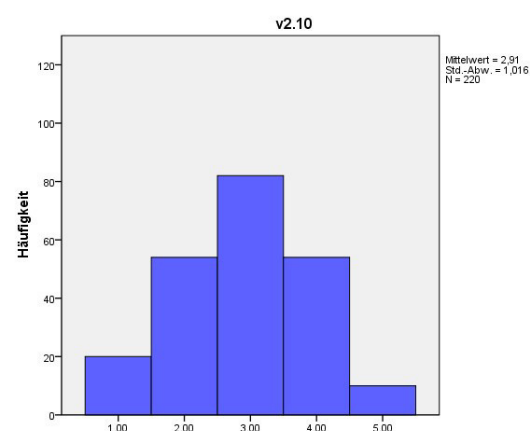


Abb. V.a10: Häufigkeitsverteilung von Item 2.10
„Das Leben in einer Umgebung, innerhalb der viele verschiedene Ethnien aufeinandertreffen, ist von andauernden Konflikten geprägt.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

Anhang V: Häufigkeitsverteilungen (Pilotstudie)

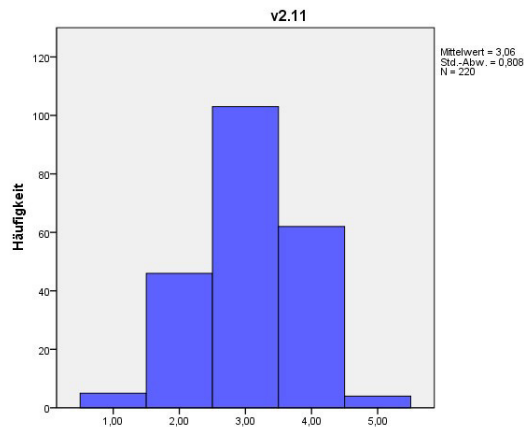


Abb. V.a11: Häufigkeitsverteilung von Item 2.11
„Von den meisten Menschen kann man annehmen, dass sie das, was sie sagen, auch tun werden.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

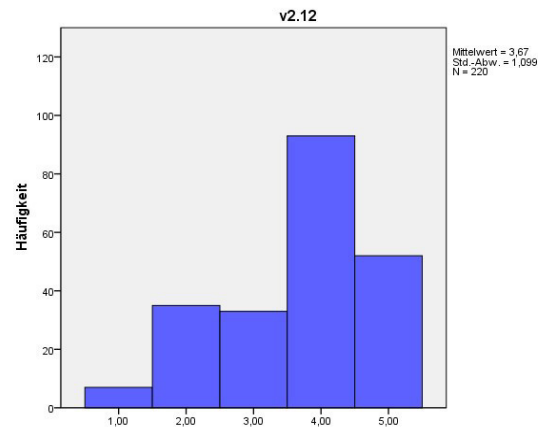


Abb. V.a12: Häufigkeitsverteilung von Item 2.12
„Es ist nicht schwer, Menschen zu finden, die einem beim Umzug helfen würden.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

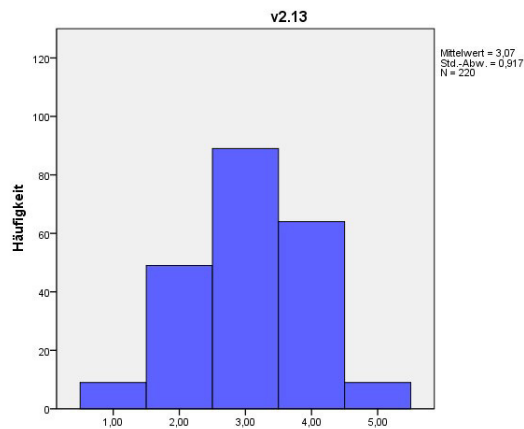


Abb. V.a13: Häufigkeitsverteilung von Item 2.13
„Der Durchschnittsmensch ist aufrichtig betroffen von den Problemen anderer Menschen.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

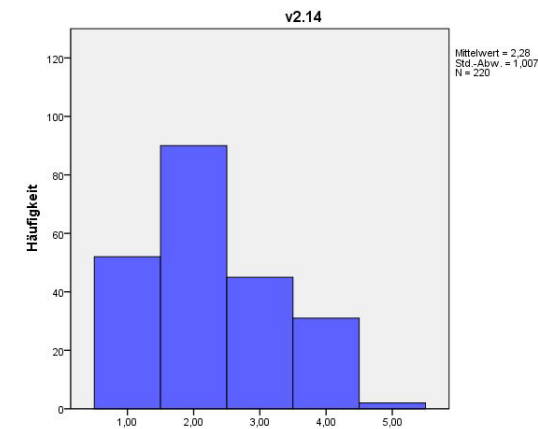


Abb. V.a14: Häufigkeitsverteilung von Item 2.14
„Es gibt nur wenige Menschen, die Leib und Leben riskieren würden, um jemandem zu helfen.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

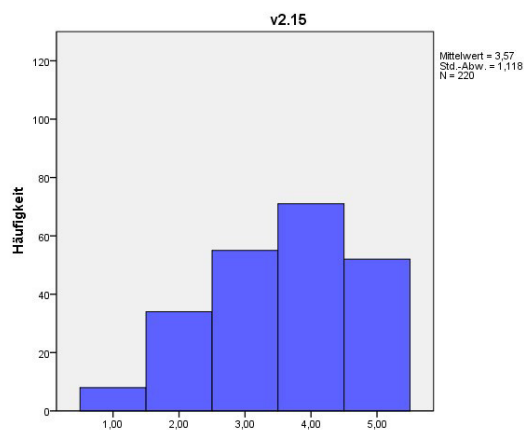


Abb. V.a15: Häufigkeitsverteilung von Item 2.15
„Am sichersten fährt man mit der Annahme, dass alle Menschen einen böartigen Charakterzug haben, der bei gegebener Gelegenheit zutage tritt.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

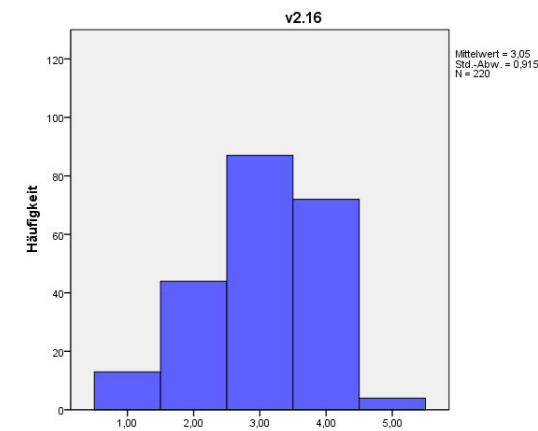


Abb. V.a16: Häufigkeitsverteilung von Item 2.16
„Hat man in irgendeiner Angelegenheit mit staatlichen Behörden zu tun, so wird man in der Regel fair und gerecht behandelt.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

Anhang V: Häufigkeitsverteilungen (Pilotstudie)

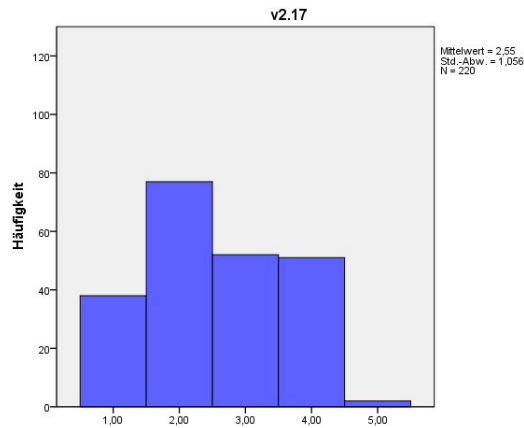


Abb. V.a17: Häufigkeitsverteilung von Item 2.17
„Die Zukunft ist zu ungewiss, als dass man weit im Voraus planen könnte.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

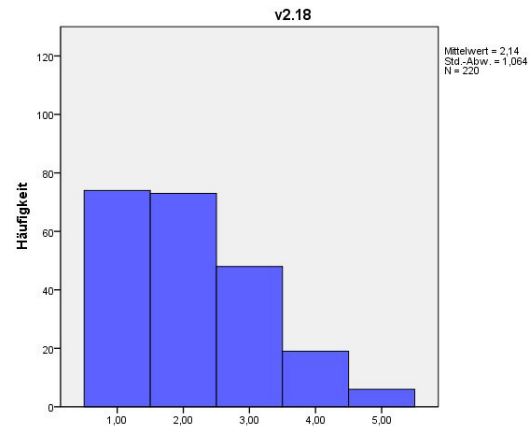


Abb. V.a18: Häufigkeitsverteilung von Item 2.18
„Der Umgang mit Personen aus dem näheren persönlichen Umfeld gestaltet sich einfacher als der Umgang mit Fremden.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

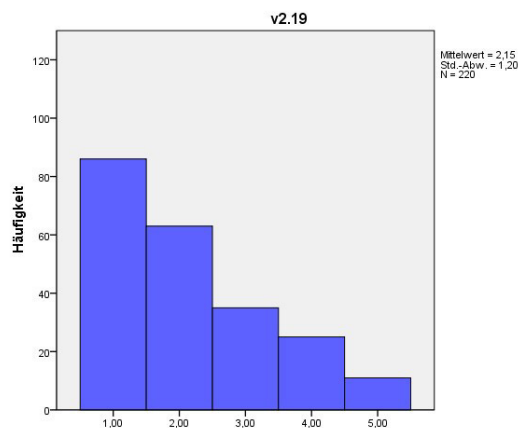


Abb. V.a19: Häufigkeitsverteilung von Item 2.19
„Es spielt keine Rolle, ob eine Wohngegend als sicher gilt, man sollte seine Wohnungstür immer abschließen, wenn man das Haus verlässt.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

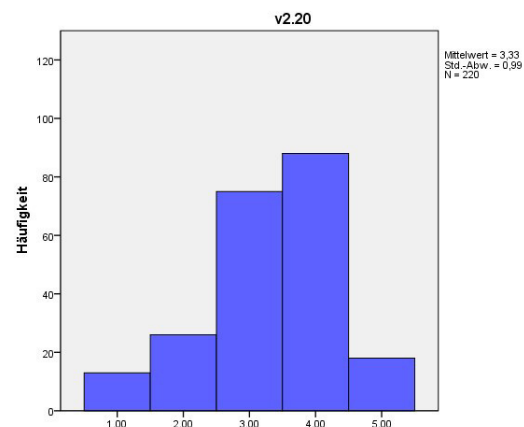


Abb. V.a20: Häufigkeitsverteilung von Item 2.20
„Der Durchschnittsmensch ist weitestgehend Herr über sein eigenes Schicksal.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

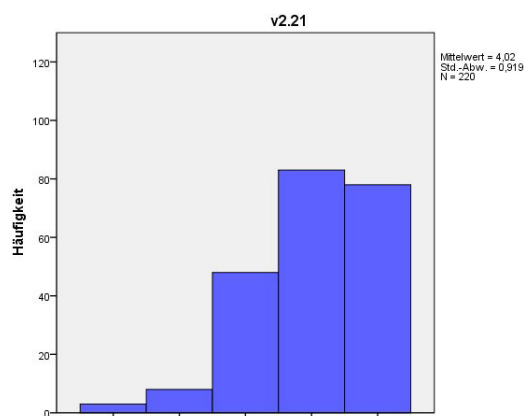


Abb. V.a21: Häufigkeitsverteilung von Item 2.21
„Der persönliche Einsatz für Vereine und Organisationen stellt eine erfüllende Freizeitbeschäftigung dar.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

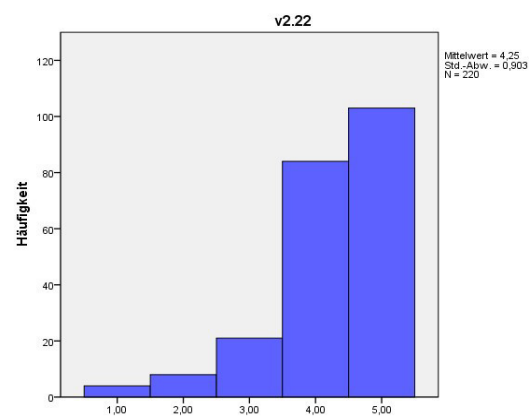


Abb. V.a22: Häufigkeitsverteilung von Item 2.22
„Erst wenn man bekommen hat, was man vom Leben erwartet, sollte man sich mit der Ungerechtigkeit in der Welt beschäftigen.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

Anhang V: Häufigkeitsverteilungen (Pilotstudie)

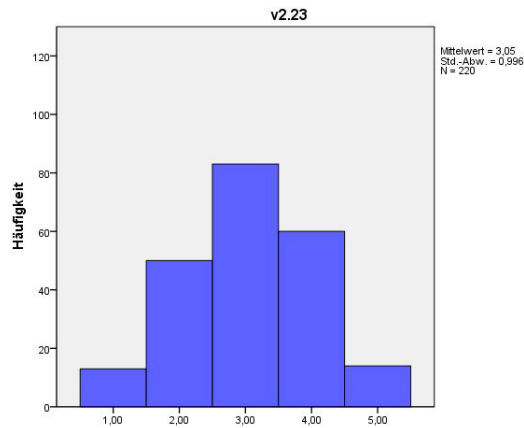


Abb. V.a23: Häufigkeitsverteilung von Item 2.23
 „Die meisten Menschen sind nicht aus Aufrichtigkeit ehrlich, sondern weil sie Angst haben erwischt zu werden.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

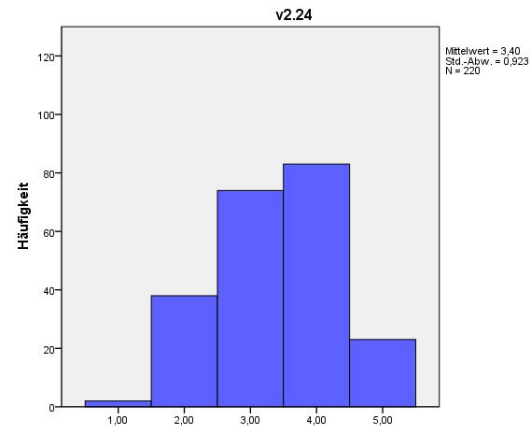


Abb. V.a24: Häufigkeitsverteilung von Item 2.24
 „Die meisten Menschen würden nicht zögern, jemandem zu helfen, der in Not ist.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

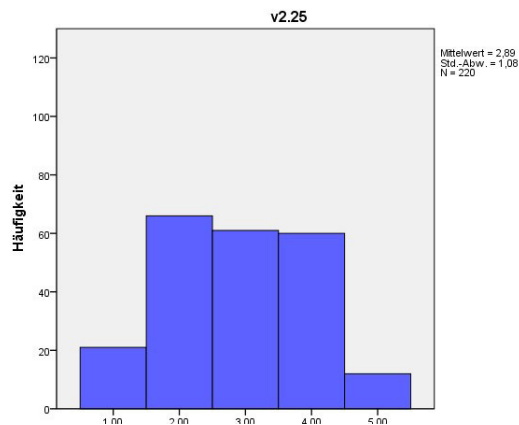


Abb. V.a25: Häufigkeitsverteilung von Item 2.25
 „Wir verlieren langsam unsere persönliche Freiheit, weil die Behörden sich immer weiter ausdehnen.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

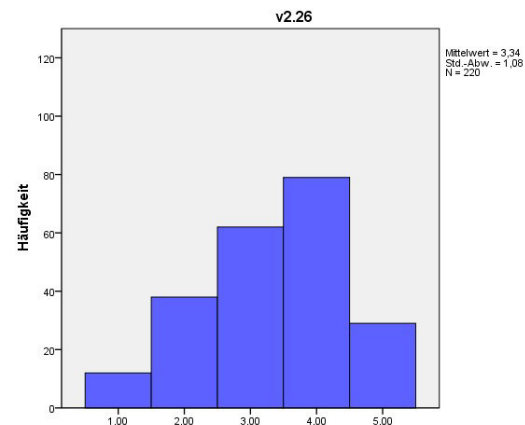


Abb. V.a26: Häufigkeitsverteilung von Item 2.26
 „Auf die Dauer gesehen, gleichen sich die guten und schlechten Dinge im Leben aus.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

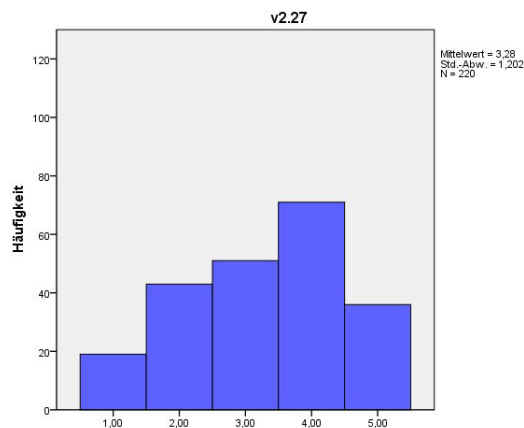


Abb. V.a27: Häufigkeitsverteilung von Item 2.27
 „Eine Kooperation von Personen, die sich ziemlich ähnlich sind, wird erfolgreicher sein, als eine Kooperation von Personen, die sich stark voneinander unterscheiden.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

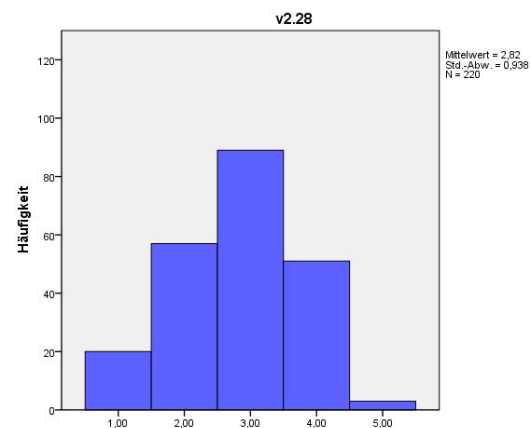


Abb. V.a28: Häufigkeitsverteilung von Item 2.28
 „Wenn man möchte, dass die Leute einen Auftrag korrekt ausführen, sollte man ihnen die Aufgabe detailliert erklären und sie genau beaufsichtigen.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

Anhang V: Häufigkeitsverteilungen (Pilotstudie)

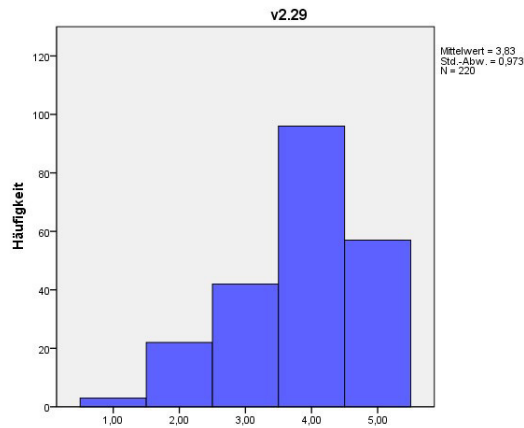


Abb. V.a29: Häufigkeitsverteilung von Item 2.29
„Man sollte nur dann den wahren Grund seiner Absichten zeigen, wenn es einem nützt.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

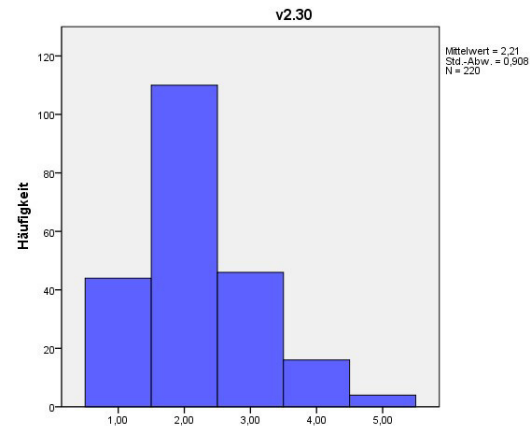


Abb. V.a30: Häufigkeitsverteilung von Item 2.30
„Die meisten Menschen sind sich gar nicht bewusst, wie sehr ihr Leben von Zufällen beeinflusst wird.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

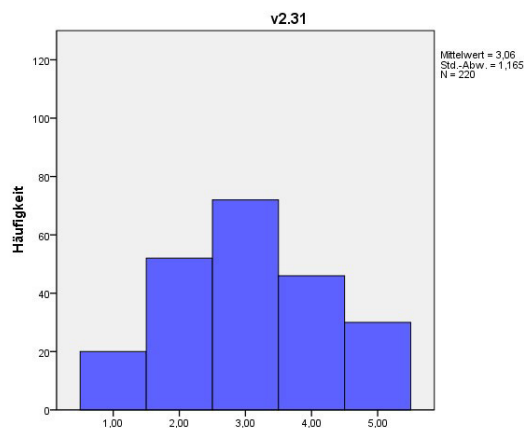


Abb. V.a31: Häufigkeitsverteilung von Item 2.31
„Das Engagement in politischen Parteien sollte innerhalb der Gesellschaft zu einem allgemeingültigen Wert werden.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

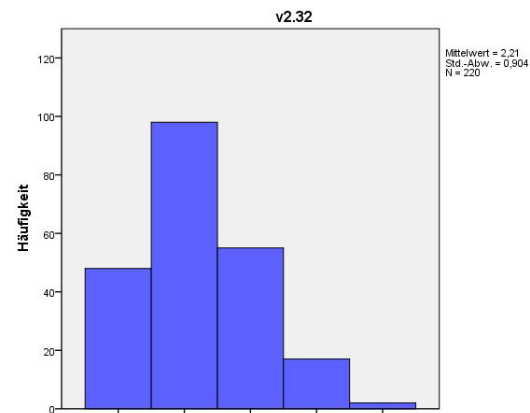


Abb. V.a32: Häufigkeitsverteilung von Item 2.32
„Die Leute geben vor, sich mehr um andere zu sorgen, als sie es tatsächlich tun.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

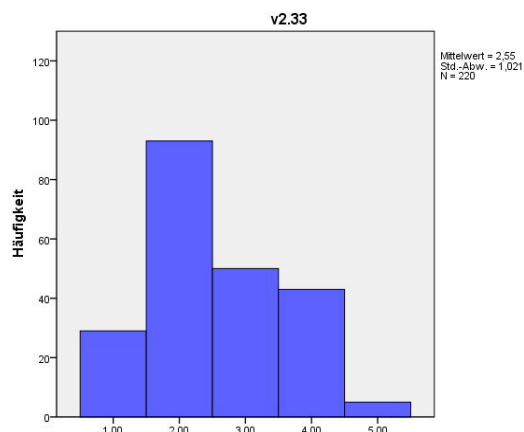


Abb. V.a33: Häufigkeitsverteilung von Item 2.33
„Es ist schwer einschätzbar, wie sich das Handeln der Menschen von Situation zu Situation verändert.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

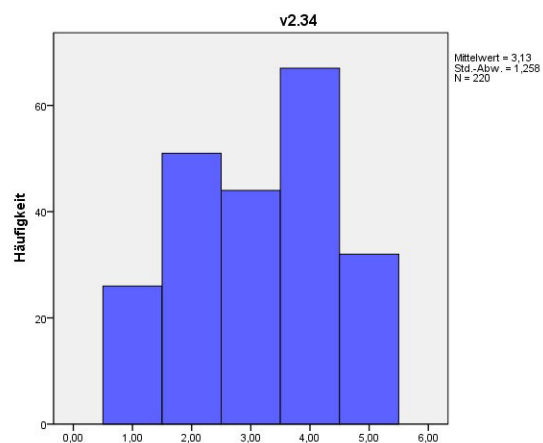


Abb. V.a34: Häufigkeitsverteilung von Item 2.34
„Es ist möglich, in jeder Hinsicht ein guter Mensch zu sein.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

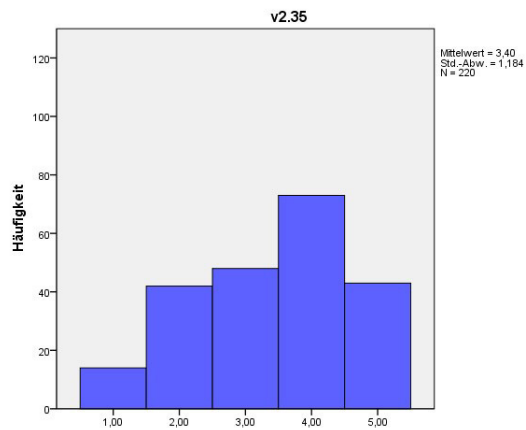


Abb. V.a35: Häufigkeitsverteilung von Item 2.35
„Der Mensch sollte nur diejenigen staatlichen Beschränkungen seiner Freiheit befolgen, deren Berechtigung er auch selber als vernünftig ansieht.“ (mit 1 = Stimme nicht zu bis 5 = Stimme nicht zu)

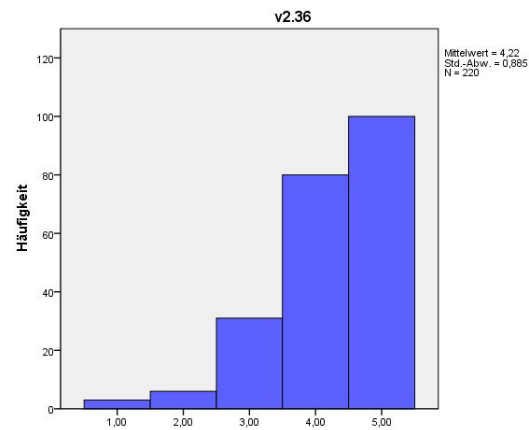


Abb. V.a36: Häufigkeitsverteilung von Item 2.36
„Man kann ein Versprechen ruhig brechen, wenn es für einen selbst vorteilhaft ist.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

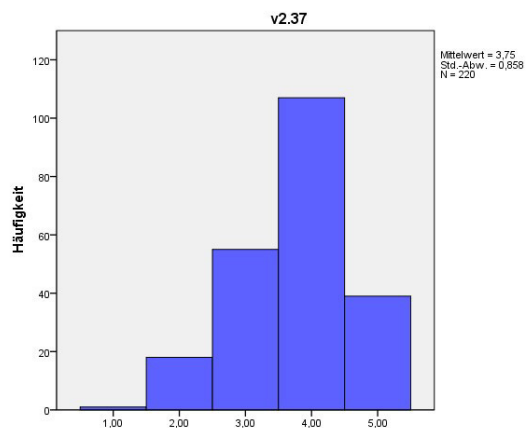


Abb. V.a37: Häufigkeitsverteilung von Item 2.37
„Wenn man anderen gegenüber in guter Absicht handelt, wird dieses Verhalten von den meisten erwidert werden.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

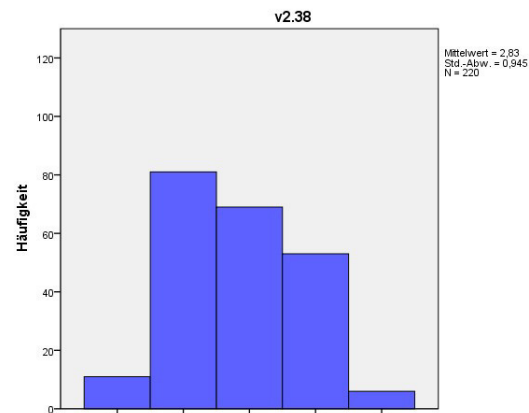


Abb. V.a38: Häufigkeitsverteilung von Item 2.38
„Die meisten Menschen haben einen guten Eindruck davon, was ihre Stärken und Schwächen sind.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

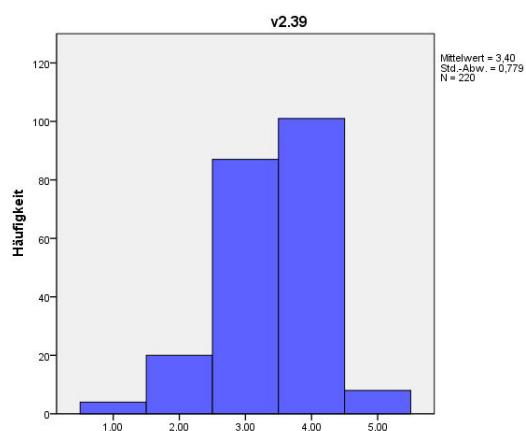


Abb. V.a39: Häufigkeitsverteilung von Item 2.39
„Wenn man einen Gegenstand an eine andere Person verleiht, wird man diesen intakt zurückbekommen.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

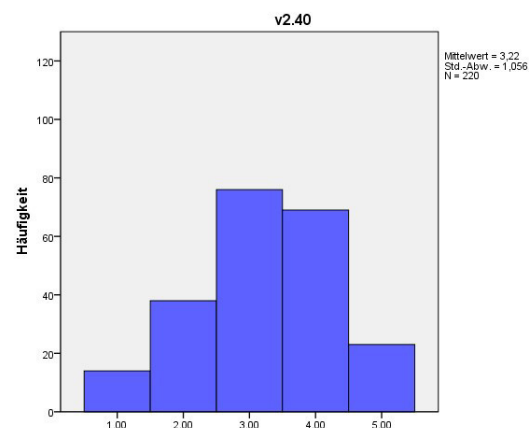


Abb. V.a40: Häufigkeitsverteilung von Item 2.40
„Heutzutage ist es nicht schwer Freunde zu finden.“ (mit 1 = Stimme nicht zu bis 5 = Stimme zu)

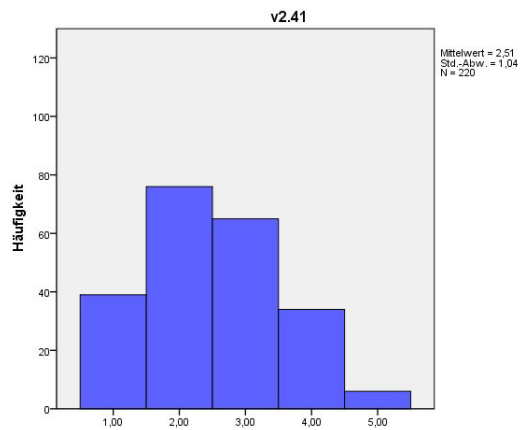


Abb. V.a41: Häufigkeitsverteilung von Item 2.41
 „Es ist besser Menschen gegenüber vorsichtig zu sein, die man gerade zum ersten Mal getroffen hat, bis man sie besser kennt.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)

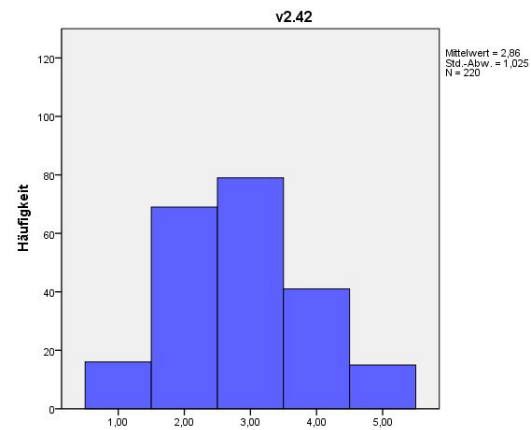


Abb. V.a42: Häufigkeitsverteilung von Item 2.42
 „Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern.“ (mit 1 = Stimme zu bis 5 = Stimme nicht zu)